DOCUMENT RESUME

ED 107 280 IR 002 024

AUTHOR Grosch, Audrey N.
TITLE Current and Retrospective Sources of Machine Readable Monograph Cataloging Records; A Study of Their Potential Cost and Utility in Automated System Development at the University of Minnesota.
INSTITUTION Minnesota Univ., Minneapolis. Univ. Libraries.
PUB DATE Mar 75
NOTE 56p.

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS Bibliographic Citations; *Cataloging; Computers; Cost Effectiveness; Data Bases; Indexes (Locaters); Information Retrieval; Information Storage; *Information Systems; *Library Automation; Library Technical Processes; *On Line Systems; Union Catalogs; *University Libraries
IDENTIFIERS File Conversion; Machine Readable Cataloging; *MARC II; Minicomputers; Twin Cities; University of Minnesota

ABSTRACT

Alternative approaches to the building of monographic bibliography files for an on-line data management system using minicomputers at the University of Minnesota Libraries' Twin Cities Campus center are described. Secondary and primary sources of the Machine-Readable Cataloging (MARC) II records are considered--including Blackwell-North America, Information Dynamics Corporation, BIBNET, and the Ohio College Library Center (OCLC)--as potential sources of retrospective and current MARC II records. File overlap comparisons and a sample of the University of Minnesota Libraries, Twin Cities Campus Union Catalog are included. In addition, methods of partial retrospective conversion and costs of using other bibliographic files in machine readable form are presented--specifically the University of Chicago Library, the University of California at Berkeley, and the New York Public Library Research Libraries files. Cost-effectiveness analyses of the various alternatives are presented. (Author/DGC)

TABLE OF CONTENTS

Page

ABSTRACT

1. Introduction and Acknowledgement 1

2. University of Minnesota Libraries and its system development plan. 2

3. General recommendations and conclusions. 3

4. MARC II monograph cataloging source alternatives and specific conclusions 4

5. Retrospective existing machine readable cataloging file usage and specific conclusions. 13

6. In-house catalog record conversion using the on-line mini-computer system. 17

APPENDICES

1. University of Minnesota union shelf list sample and application of the sample to the total union catalog, including MARC II file overlap 20

2. Detailed findings for MARC II monograph cataloging sources as found in Section 4. 25

3. Detailed findings for retrospective existing machine readable cataloging files as found in Section 5. 38

4. Supplementary information concerning in-house catalog record conversion as found in Section 6. 47

ABSTRACT

This report discusses various alternatives and their costs for building monographic bibliographic files for an on-line data management system using minicomputers which is under development for the University of Minnesota Libraries, Twin Cities Campus. Secondary and primary sources of MARC II records are considered, including BLACKWELL-North America, Information Dynamics Corp. BIBNET and Ohio College Library Center (OCLC) as potential sources of retrospective and current MARC II records. File overlap comparisons and a sample of the University of Minnesota Libraries, Twin Cities Campus Union Catalog are included. In addition methods of partial retrospective conversion and the costs of using other bibliographic files in machine readable form are presented - specifically the University of Chicago Library, the University of California, Berkeley and the New York Public Library Research Libraries files. In-house conversion costs on the on-line minicomputer system are presented as derived on the system installed in the University's Bio-Medical Library. The findings support building and storing at least a partial MARC II file on-line with the remainder on removable disc packs as a less costly alternative to telecommunication transmission of MARC II data from BIBNET or OCLC at their current subscriber costs. For retrospective conversion, the costs to convert in-house directly from catalog cards using the on-line minicomputer system are lower than those obtainable via edited use of the three large research library files mentioned above.

1. Introduction and Acknowledgement

The purpose of this study is to investigate the utility, availability, and costs of various sources of current MARC II monograph cataloging records as well as certain large retrospective monograph catalog files for potential input into the University of Minnesota Libraries on-line mini-computer data management system now under development.

For current cataloging needs for monographs a library system should secure Library of Congress MARC II cataloging information. A number of options are open from which a library can choose. This choice will be determined by acquisitions costs, operational cost, and expected volume of records to be used as well as the type of system the library has in which the records will reside.

Retrospective records are particularly useful for implementation of circulation control or on-line terminal access to the active collection. There are also several procurement options available for a large academic research library. The choice of which of these files or if any existing file should be used can be made by determining the potential number of useful records, the costs to acquire and process the data, and the cleanliness/completeness of the cataloging data. From this information per record costs may be derived and compared with local catalog record conversion costs to determine the cheapest and highest quality procedures.

Serial data records are not included in this study as the Minnesota Union List of Serials (MULS) contains over 71,000 MARC II serials format records, including all known University of Minnesota serial holdings.

Since many of the conclusions on file utility depend upon determining the potential useful number of records in a file, a proportional random sample of titles was extracted from the various shelf-lists comprising the cards contained in the University of Minnesota Libraries Union Card Catalog. The validity of this sample was determined by comparing its characteristics with known reported characteristics during the last 10 years of catalog department reporting. A further practical test was made by comparison with a large retrospective catalog data base and then using a similar sample of pages from this large retrospective catalog checking all entries on those pages in the University of Minnesota Catalog. A +1% variance was found, confirming via a practical test that the overlap figures determined will be conservative but highly accurate. At the same time MARC potential was also derived so that one could determine how many retrospective and current MARC II records would be useful. A description of this sample is included in the Appendix.

Section 2. describes the system development plan of the library. Section 3. gives general recommendations and conclusions of this study. Section 4. gives the specific conclusion for MARC II file access among the six alternatives considered. Details of the study of these alternatives are included in the Appendix. Section 5. gives the findings regarding use of three existing library system machine readable data bases - the University of Chicago, New York Public Library, and University of California files. The Appendix contains specific information on these source files, their characteristics and costs. Section 6. describes the methodology and cost of in-house catalog record conversion using the on-line minicomputer system.

It is our belief that the information presented here will be useful to other libraries in either its methodology to conduct their own study related to their cataloging needs or at least narrow the cost factors to make better choices of the alternatives as libraries proceed with their individual automation programs.

The author of this report would like to extend deep appreciation to each of the commercial firms who gave information on their files. Northwestern University Library provided assistance in the comparison of the University of California, Berkeley Five Year Union Catalog through their Reference Department. The Smithsonian Institution Library and in particular Mr. Philip Leslie, Assistant Director provided invaluable assistance to access the OCLC file. Special mention is due Mr. Don Norris of our own Research and Development Department for providing the estimates for constructing a MARC II tape file from the raw data tapes and Ms. Elizabeth Lange, Head of our Catalog Division for assistance in the shelf list sample extraction. The staff of the Research and Development Department and Mr. Glenn Brudvig, Assistant Director for Research and Development served as reviewers of this report. For their critical comments this author is indebted.

2. University of Minnesota Libraries and Its System Development Plan

To understand certain estimates and the following conclusions in this report the reader will find a brief view of the University Libraries and its system plans helpful.

The University Libraries on the Twin Cities Campus are composed of multiple service units including-subject special libraries ranging in size from 20,000 to 240,000 volumes. The largest service unit, Wilson Library, houses the major active portions of the general book collection as well as serving as system headquarters. Except for the Bio-Medical Library and the St. Paul Campus Library all technical processing functions are performed centrally in Wilson Library. The Union Catalog for the campus is maintained there. The collections total 3.5 million volumes, with book circulation of over 1.8 million transactions per year for the system. In addition the University Libraries serve as a statewide resource and provide loan and photocopy to all libraries in the state as well as neighboring states through network arrangements. Also various subject network activities are currently supported or under development.

The system development plan for the University of Minnesota Libraries involves the creation of an on-line integrated data mangement system capable of supporting the traditional technical processing and reference service activities in a large library. Acquisitions, accounting, cataloging, serials management, circulation, and bibliographic searching generally comprise these activities. In addition the system is being conceived as a dynamic library management tool to provide data reduction and analysis for those concerned with the library's management. The system will eventually enable tying existing on-line terminals to virtually any on-line information retrieval data base for subject searching and linking of the search results to our own collection resources. Moreover, the system may also be used as a message storage and forwarding system to support various network activities such as document delivery, collections coordination, and bibliographic search facilities.

This system is being built using dedicated minicomputer systems (Digital Equipment Corp. PDP 11 series processors) using advanced software techniques, and peripheral equipment from a variety of special vendors. Eventually the system will comprise a series of such computers linked together into a network, with subsystems to perform specific functions pertinent to a single library or a group of libraries. This system will permit individual libraries a degree of customization in their application needs yet enable common software to be used throughout the system's parts. The Bio-Medical Library node in this network is the initial one now being installed. Planning is underway to move forward with similar systems in the Wilson Library and St. Paul Campus Library. Other units and applications beyond those of traditional operational technical processing would be added soon as possible in the future.

The major benefits of this system approach are:

1. Lower hardware and operation costs,

2. Greater modularity within the system to incorporate new applications, hardware, and software without affecting the total system operation,

3. Common software maintenance and enhancement, yet the ability to tailor a data base and input/output portions of a system to the specific service unit,

4. Lower cost hardware maintenance and enhancement, including replacement or expansion.

4. Lower cost hardware redundancy in case of system failure,

5. Modular installation and system evolution without the large dollar investment required to acquire dedicated large central computer support,

6. Control over total system environment,

7. Problem minimization when dealing with large file systems,

8. Compatability with long term plans of the library due to high modularity and library control over the system.

One can readily see the need for source bibliographic data in such a system. It is hoped this study will provide information helpful in the planning of systems of this type, both at the University of Minnesota and in similar libraries.

3. General Recommendations and Conclusions

There are many facets to the study of the utility of machine readable files contructed by other institutions. This study has attempted to address the use of certain specific files using a methodology of first determining the number of applicable records which could be obtained from the file, then considering the quality of the records and their cost to procure and process. Their procurement and processing has been viewed within the

the context of the on-line dedicated minicomputer system rather than in some other type of system.

The following conclusions resulted from this study:

- The University of Minnesota Libraries should procure an already cumulative retrospective MARC II file and continue updating it with the weekly MARC II ALL LANGUAGE service.

- The Hennepin County Library cumulative MARC II file, of the alternatives examined, or another no fee source would be the least costly method of obtaining this file.

- The University Library's own system would be a cost effective storage site for on-line MARC records of several years recency.

- With changes in the number of applicable MARC records to Minnesota tending to increase as further languages are added it will be increasingly cost effective to maintain an on-line MARC II file.

- The University Library's system, through usage monitoring of the 500,000 record present full MARC II file could determine just how much of the total file need be system resident and how much could be made system resident merely on a scheduled basis.

- If the MARC II file usage on the University's system were to expand because of other library's usage of the file it may be cost effective to maintain a full MARC II on-line file.

- Commercial or other sources of MARC II records considered in the study cannot provide less costly source MARC II records than the Hennepin County file and our processing of that file except if:

    1. on-line computer-to-computer high speed station-to-station telecommunications were used, and

    2. the record charge rate was to be that now charged by L.C. for the weekly MARC II subscription, i.e. approximately $.033 per record, or $.068 per applicable record.

- Transmission of a MARC II record via 4800 BAUD dial up-station-to-station method from the BIBNET data base is approximately the same cost as the per record cost of the MARC II subscription itself, i.e. $.032 per record.

- Computer-to-computer transmission costs and record usage charges would be cost effective if they were somewhat higher than local MARC II tape service procurement and processing due to the relief from planning additional disc storage space locally for records with low user potential. An exact dollar figure is difficult to determine and would have to be considered in light of the disc costs and all other factors known at the time such a link was to be investigated in greater detail than possible for this study.

- The use of retrospective non-MARC records in the University's system would apply to creating files for circulation control and on-line access to the active portions of the collection at this time.

- From the shelf list sample and some rudimentary knowledge of the circulation of books in the library it appears that using a publication date cutoff of 1960-to date would produce such an active title file in the system. Its size is estimated at 377,500 titles.

- Conversion of non-MARC retrospective titles by in-house keyboarding on CRT terminals connected to the library's own minicomputer, working directly from a catalog main entry card would cost $.81 per record, i.e. slightly under the cost per initial use of an OCLC record including producing cards.

- Use of retrospective files such as the University of Chicago, New York Public Library, or University of California would produce a higher total cost as acquisition, programming, and temporary storage costs of these files must be added, with only an estimated 30% reduction in human labor through the use of these records which will require individual human editing.

- File quality improvements cannot be guaranteed through use of the above retrospective files as their quality varies greatly and no comparison of their quality with the quality of our comparable records has been made.

4. <u>MARC II monograph cataloging sources alternatives and specific conclusions</u>.

Six alternatives for acquiring MARC II monograph cataloging information pertinent to the University of Minnesota Library's collections were examined. Table 1. shows the costs of each of these alternatives over a five year period and an average annual cost over that period. In addition per record costs are also given.

Alternatives 5 and 6 which specify direct station-to-station 4800 BAUD transmission of data from the Information Dynamics Corp. BIBNET and Ohio College Library Center files respectively do not include any of their site programming costs. We have further assumed that their current record usage charges would also apply since such a service does not now actually exist in this form. There is merit in considering direct computer-to-computer transmission of batched search requests and batched output requests via a dial-up station-to-station switched line operating at 4800 BAUD.

Examination of Alternative 3 which provides for local support of a full MARC II file with file size reduction by 4/5 ths within two years and an equivalent of 2 years of MARC II records resident on-line reveals the lowest total cost over an annual and five year period. Although per record cost is higher, every record is a used record as opposed to the high used record costs in Alternatives 1 and 2.

Alternative 3 is the most cost effective way of presently providing MARC II data over the planned system at this time. If conditions of record usage in the future indicated that a higher percentage of MARC II records would actually be utilized in the system, then Alternative 2 would probably approach the per applicable record costs of the present preferred Alternative 3.

As the utility of remote transmission of MARC II records on a cost basis is dependent upon the size of the per record usage charge and the use of dial-access switched 4800 BAUD transmission it remains to be seen whether agencies such as Information Dynamics and the Ohio College Library Center can offer this service on a more cost effective basis than their current charges would permit. Further study would be required to cost any such announced services as they became available to the University of Minnesota. This study would require costing of the disc storage savings as well as the system and programming factors.

The detailed findings, cost estimates, and components studied to arrive at these alternatives and specific conclusions maybe found in the Appendix to this report.

TABLE 1. ALTERNATIVE METHODS OF PROCESSING MARC II DATA OVER A 5 YEAR EXPECTED SYSTEM LIFE PERIOD. COSTS DERIVED FROM VALUES IN TABLE 2.

Alternative 1.

This alternative assumes building and maintaining one's own file on the PDP 11 system - a full MARC II monograph file.

Component	Total Cost	5 Years Cost	Cost Per Year
Build own cumulative file	$19,936	$19,936	$ 3,988
Current MARC II Subscription	$10,000	$10,000	$ 2,000
Updating resident file	$25,440	$25,440	$ 5,088
Residency cost of disc storage	$163,000	$163,000	$32,600
	$275,976	$275,976	$55,196

Per record cost of $1.10 (used records)
Per record cost of $.34 (all expected 800,000 records)

Alternative 2.

This alternative assumes acquiring a cumulative file from Hennepin County Library or another source without royalty or usage fees and maintaining the full file resident on-line to the PDP 11 computer system.

Component	Total Cost	5 Years Cost	Cost Per year
Acquiring Hennepin County File	$ 2,096	$ 2,096	$ 419
Current MARC II Subscription	$ 10,000	$ 10,000	$ 2,000
Updating Resident file	$ 25,440	$ 25,440	$ 5,088
Residency Cost of Disk Storage	$163,000	$163,000	$32,600
Hardware Maintenance	$ 57,600	$ 57,600	$11,520
	$258,136	$258,136	$51,627

Per record cost of $1.03 (used Records)

Per Record cost of $.32 (all expected 800,000 records)

Alternative 3.

This alternative assumes local support of an initial full MARC II file and then reduction in the size of the file by 4/5ths from 1968-1974 data within two years and an equivalent of 2 years worth of MARC II data on the system thereafter, i.e. a permanent file of about 154 million bytes or 2 RJP04 disk units. Remaining records would be stored off-line.

Component	Total Cost	5 Years Cost	Cost Per Year
Acquire Hennepin County file	$ 2,096	$ 2,096	$ 419
Current MARC II subscription	$ 10,000	$ 10,000	$ 2,000
Updating Resident file	$ 25,440	$ 25,440	$ 5,088
Residency Cost of disk storage	$ 138,000* (77,600)	$ 77,600	$ 15,520
Hardware Maintenance	$ 33,600	$ 33,600	$ 6,720
	$ 209,136 (148,736)	$148,736	$ 29,747

Per record cost of $.59

*Under this alternative $75,000 of disc would be relocated after 2 years to storage of the library's own data files therefore a pro-rated value of $27,600 per year for disc for the initial two years would apply with 2 RJP04 units dedicated the remaining 3 years for $50,000.

Alternative 4.

Purchase applicable MARC II record form Blackwell-North America for retrospective and current MARC II records and then process on our system.

Component	Total Cost	5 Year Cost	Cost Per Year (over 5 Years)
Blackwell File Acquisition/Prog.	$ 49,500	$ 49,500	$ 9,900
Annual value of L.C. Card No. searches	$ 85,000	$ 85,000	$ 17,000
Updating our file	$ 25,440	$ 25,440	$ 5,088
Disc storage costs for 284,090 records for 5 years = 3RJP04	$ 88,000	$ 88,000	$ 17,600
Hardware maintenance	$ 36,000	$ 36,000	$ 7,200
	$283,940	$283,940	$ 56,788

Per record cost of $1.00

Alternative 5.

Secure from BIBNET via cheapest way MARC II records which match University of Minnesota titles and process on PDP 11/40 system.

Component	Total Cost	5 Year Cost	Cost Per Year (over 5 years)
Batch acquisition of 100,000 retrospective MARC records	$ 90,000	$ 90,000	$ 18,000
Current acquisition of 5 years of MARC records usage fee for 150,000 records @.90 each	$135,000	$135,000	$ 27,000
Long distance LDX® 4800 BAUD station-to-station communication	$ 5,040	$ 5,040	$ 1,008
Updating our file	$ 25,440	$ 25,440	$ 5,088
Disc storage costs for 250,000 records for 5 years, i.e. 3 RJP04 disc units	$ 88,000	$ 88,000	$ 17,600
Hardware maintenance	$ 36,000	$ 36,000	$ 7,200
	$379,480	$379,480	$ 75,896

Per record cost $1.51

Alternative 6.

Secure from OCLC via cheapest way all MARC II and shared cataloging file records which match the University of Minnesota's holdings and process on PDP 11/40 system.

Component	Total Cost	5 Year Cost	Cost Per Year
Usage charges for 265,200 matching retrospective records at $.904 each	$239,741	$239,741	$ 47,948
Current cataloging volume of 39,336 each per year at $.904 each for 5 years.	$177,798	$177,798	$35,560
Cheapest method of on-line communications-dial up LDX 4800 BAUD station-to-station for total 461,680 records	$ 14,774	$ 14,774	$ 2,955
Disc storage 324 million bytes equals 4 RJP04 units	$113,000	$113,000	$ 22,600
Hardware maintenance	$ 48,000	$ 48,000	$ 9,600
	$593,313	$593,313	$118,663

Per record cost of $1.28

TABLE 2. COMPONENT COSTS WITHIN ALTERNATIVE METHODS OF PROCURING MARC II MONOGRAPH RECORDS SHOWN IN TABLE 1. (Detail discussion in the Appendix to this report).

Cost Component	Total Cost	Per Record Cost	Per Usable record cost
Build own cumulative MARC II file.	$ 19,936	.04	.20
Annual all language MARC II subscription (60,000 annual titles 29,386 usable)	2,000	.033	.068
Update PDP 11/40 system resident MARC II file.			
Initial year (incl. $3600 programming)	7,968	.13	.27
Subsequent years	4,368	.07	.15
Residency cost of MARC II file on PDP 11/40 system - RJP04 disk units.	113,000	.23	1.13
Equipment maintenance	9,600	.02	.096
Annual MARC II file growth cost per year (1/2 RJP04 equiv. per year).	13,000	.23	.44
Added annual equipment maintenance	2,400	.04	.08
Acquire Hennepin County MARC II Cum. File (or other file at no royalty fee)	2,096	.004	.02
Programming Henn. Co. File Conversion	3,600	.0072	.012
Loading file onto PDP 11/40 system discs	280	.0006	.003
Subtotal	5,976	.012	.035
Blackwell-North American (ABEL) File Purchase	45,000	.073	.274
individual per record costs for selected records est. 164,096 records	16,409 - 41,023		.10 - .25
Annual fees L.C. Card No. Searches	17,000		.34
Programming Blackwell file conversion	4,500	.01	.03
Loading file onto PDP 11/40 system discs	280	.0006	.003
Information Dynamics BIBNET File fee off-line 100,000 records	90,000		.90
BIBNET 2707 on-line 300 BAUD service (100,000 records)	209,475		2.09
BIBNET 2707 Annual (29,336 record est.)	14,250		.48
LDX station-to-station 4800 BAUD transmission only for 100,000 records	3,276		.032
LDX station-to-station 4800 BAUD transmission for search and send 30,000 rec.	1,008		.032
Communication LDX 4800 BAUD annual plus per record usage charge	29,400		.98

OCLC MARC II/shared catalog file retrospective costs of est. 265,200 rec.	526,262		1.98
OCLC Annual costs for 39,336 rec. without cards - 4800 BAUD private line	44,272		1.125
OCLC Terminal/card production services on est. 39,336 rec. yr. incl. card print.	96,925		2.15
OCLC - PDP 11/40 private line 4800 BAUD communications costs 265,200 rec.	2,175		.0082
OCLC Annual costs for 39,336 rec. without cards via 9600 BAUD comm to PDP 11/40.	45,160		1.148
OCLC-PDP 11/40 LDX 4800 BAUD station-to-station annual 39,336 rec.	37,720		.958

5. Retrospective existing machine readable cataloging file usage and specific conclusions.

There are many existing files of machine readable monograph cataloging records. Some of these files are maintained by commercial firms along with MARC II file information. An example of such a file is the Blackwell-North American offering. Other files have been constructed by libraries either through a vendor specializing in catalog preparation or through their own system creation efforts.

As the University of Minnesota collections represent those of a large academic research library we have chosen to compare files of a similar nature as large files of this type would tend to have a larger number of potentially useful records. Therefore, we have considered the following source files, even though two of these files would contain MARC II derived cataloging records as well:

1. New York Public Library. Research Libraries catalog data base (including MARC II records),

2. University of Chicago Library catalog retrospective data base (including MARC II records),

3. University of California, Berkeley. Five year union catalog supplement data base.

4. Ohio College Library Center (OCLC) shared catalog records.

Table 3. Comparison of costs of retrospective existing machine readable cataloging files.

File Name	Usable Total Records	MARC II	NON-MARC II	Total Cost	Cost Less Acquisitions Cost	Total NON-MARC Per Record Cost	Total All Record Costs	Total NON-MARC Less Acquisitions Per Rec. Cost	TOTAL All Records Less Acquisitions Per Record Cost
New York Public Library Research Libraries	63,338	31,036	32,302	55,382	33,725	1.71	1.05	1.04	.53
Univ. of California Berkeley. Five Year Union Catalog	232,500	-	232,500	195,888	115,888	.84	.84	.49	.49
University of Chicago Library Catalog	95,000	48,450	46,550	60,325	33,825	1.29	.63	.73	.36
OCLC Shared Catalog	165,200	-	165,200	327,096	-	1.98	-	-	-

Table 3. shows the potentially usable records contained in each of the above files together with the number of MARC and non-MARC II records. The total cost of each file has been derived according to detailed assumptions and calculations found in the Appendix. As in some cases an acquisition cost could not readily be determined but only conjectured, the cost exclusive of acquisition costs has been shown. Further, per record costs under a variety of arrangements have been calculated.

From this table it can be seen that the University of California, Berkeley file produces the largest extension of titles beyond those covered by MARC at a cost of $.84 total based on a usable potential of 232,500 records out of over 750,000 records in the full file. However, the quality level of the records and the lack of easily extracting the L.C. card No. for searching do present problems in the use of this file.

At the other side, the Ohio College Library Center shared catalog would produce potentially 165,200 records at a cost of $1.98 exclusive of costs local to the University of Minnesota. These records appear to be high quality but at a large cost - in fact about the same cost as is shown in Section 6. for in-house catalog record conversion using the on-line minicomputer system. The New York Public Library and University of Chicago Library files are in between these two alternatives.

Considering these costs to procure, program, temporarily store, and edit these files must be weighed against the conversion of existing in-house cataloging records using MARC II records where existing and direct data conversion via terminal for the remainder. A large share of cost in use of another file is the amount of editing of records necessary, either to correct errors or bring them into agreement with the locally produced catalog card in hand.

In Section 6. it has been estimated that approximately $.20 per record input could be saved by having a high quality pre-machine readable record available for editing or augmentation. The cost to acquire, program and temporarily store these data files considered here do not offset the conversion costs in-house. In fact, for the New York Public Library file, at the above costs Minnesota would produce $6,460 saved on processing 32,302 records, but would have expended $33,725 minimum to effect this saving.

Similarly, the University of California, Berkeley file would save $46,500 in processing 232,500 records but in so doing require an expenditure of $115,888 minimum. Even, the University of Chicago file shows a similar condition. For a saving of $9,310 or 46,550 records it would be necessary to expend a minimum of $33,825. For the OCLC shared cataloging file the University would save $32,040 on processing 165,200 records but expend $327,096 just to acquire these records.

Therefore a net loss of $27,265, $69,388, $24,515, and $294,056 respectively occur if these files are used.

Here-to-fore many librarians have assumed that the use of a machine readable record file created by others should automatically be used because cost savings will result together with perhaps potentially better quality records. But this study shows that such is not the case. Even verified MARC II records do require occasional alteration or addition of locally

specific information, but here the record is obtained cheaply, in a standard form, and with uniform identification of its data elements together with the credability level of cataloging made machine readable under close control.

The reasons for these net losses above can be summarized as follows:

1. The processing required to support on-line editing of a foreign file can involve more programming and temporary disc storage facilities than use of standardized MARC II file handling program modules and present standard system editing routines.

2. The percentage of usable records in a file will probably not exceed 40% - 50% for any monograph file causing support to an extensive searching system to retrieve the record required, both by L.C. card numbers or author/title searching.

3. The quality of the source file may be questionable and in some cases the addition or correction determination and action takes as long in labor as entering the original record without this search or determination.

4. Only about 30% of data entry personnel time can be saved per record manipulated in machine form as the record still must be checked and local data entered. This amounts to approximately $.20 per record labor at the University of Minnesota.

5. If human alteration of records manually would not be required, then total automatic record selection from several files could be performed via computer and a merged file created cost effectively. However, even with good format recognition techniques, a high quality bibliographic record will require human attention.

6. The use of a dedicated minicomputer system brings a lower systems environment cost than if comparable support were costed on a large system, making the overall costs to be offset even more pronounced.

7. The relatively low cost of L.C. MARC II information together with an accepted level of quality permits the best potential economies in pre-machine readable records procured. Until larger numbers of retrospective records are available on these terms the costs for retrospective file conversion will be somewhat higher than those for creating a current cataloging record from MARC II source information.

Therefore, at the present time retrospective conversion of titles which are non-MARC II appears to be most cost effectively accomplished by conversion of existing catalog entries directly from the main entry cards. At least the same record quality level as found in the original file should result if appropriate quality editing is performed. Typographic quality should be excellent. Cataloging inconsistencies should duplicate those of the manual catalog but be easier to identify via the system, and subsequently easier to correct.

Ultimately computer based authority file controls as found in the New York Public Library system will result in a file of maximum consistency when built over the long time of a library catalog's life.

6. In-house catalog record conversion using the on-line minicomputer system.

The previous section of this report has shown that costs to use files created elsewhere have a negative cost effect. This effect has been determined by working out a probable methodology to convert active portions of the University of Minnesota union catalog using the on-line minicomputer system for data entry, modification, proofing and creation of this converted file.

With an estimated 1.25 million titles cataloged within the University of Minnesota Libraries, our shelf list sample has produced some specific indications of the publication dates of titles. In the Appendix the Table of Sample characteristics shows a total of 377,500 titles having imprint dates of 1960 or later. The University's Bio-Medical Library is now initiating a file conversion of titles on its on-line minicomputer system within that period. Of these titles, it appears from the sample that 100,000 will be available MARC II records.

In the recommended method described here partial conversion of the greater portion of the active collection would occur. This method consists of selecting main entry cards from the union catalog for conversion by publication date and the physical conversion process itself. Various alternatives have been considered for the selection of main entry cards, but the lowest cost alternative is the one which involves the least human secondary file handling, decision making, or temporary work record creation. A direct search of the catalog to extract main entry cards by date of publication is least costly. The physical conversion process itself then involves entering the data elements from the source record directly using a CRT terminal. If the record is a MARC II record or suspected to be one, a search of the temporarily resident full MARC II file would be made. A record so found would then be completed with the local information such as call no. recorded. If the record was not on the MARC II file then a complete keyboarding would be required. Editing, proofreading, correcting the records by another operator on a CRT terminal would complete the process. Other cost components would be personnel training, documentation, terminal/communications, and software costs. A resident MARC II file as described earlier in Section 4 of this report is assumed available for use. Table 4. gives the cost details for this method. The method assumes a 3 year time period for the conversion effort to be accomplished.

It cannot be determined exactly how much time will be saved by modifying an existing record in machine readable form. It is conservative to conclude that only about 30% labor on input would be saved since the record still must have local data entered and must be checked, proofread, and perhaps corrected. Such a labor saving would lower the input data entry portions of conversion costs by $.20 per record so manipulated. Use of a MARC file already resident for current cataloging purposes would save approximately $20,000 in labor on this basis.

Table 5. gives the labor rates used to produce the cost information in Table 4.

The Appendix to this report contains some additional detailed information which has led to the recommended method and costs discussed in Section 6.

Table 4. Costs of in-house conversion of 377,500 monograph cataloging records using the on-line minicomputer system.

Operation Selection	Cost	Per Title Cost	Total Cost
Select main entry cards by search (per 200)	$ 5.40		
Xerox labor (per 200)	2.70		
Xerox copying (per 200)	2.04		
Total cost per 200 selected	10.14		
Total costs for selection 377,500		.051	$ 19,139
Personnel training/documentation			
Project Director 6 mo. @ $1,600 mo.	9,600		
Documentation preparation	1,000		
Personnel training 16 CRT operators	18,720		
and 4 editors each 6 weeks	6,756		
Total personnel/documentation costs		.095	36,076
Supervision, editorial, error correction			
Project director 30 mo @$1,600 mo.	48,000		
Editors 4 @ 30 mo each @$1,126 mo	135,120		
CRT operators proofreading 8@30 mo. each @ $780 per mo.	187,200		
Total supervision, editorial, error correcting		.98	370,320
Terminal data entry			
CRT operators 8 @30 mo. each @ $780 mo.	187,200		
16 CRT Terminals (Super Bee SB-1 or equivalent and communications equip.)	62,800		
Equipment Maintenance 30 mo.	5,000		
Total terminal data entry		.68	250,000
TOTALS		1.806	680,535

Table 5. Labor rates used in retrospective file conversion costing and production estimates.

Function	(inc/fringe benefits) Hourly Rate	Monthly	Annual
Selecting titles, xeroxing main entry cards	2.70	468	5,616
Project Director	9.23	1,600	19,200
Editors	6.50	1,126	13,520
CRT Terminal Operators	4.50	780	9,360

Production Estimates

Editorial checking main entry cards 150 per day average	37,500/year	3,125/month
Input CRT terminal rate 75 titles per day average	18,750/year	1,562/month
Proofreading CRT terminal rate 75 titles per day average as above		

Therefore 377,500 records processed on a one shift basis would require 30 calendar months using 16 CRT Terminal Operators, 4 Editors and one Project Director.

APPENDICES

Appendix 1. University of Minnesota Union Shelf List Sample and Application of the Sample to the Total Union Catalog Including MARC II File Overlap.

University of Minnesota Libraries Shelf List Sample

The union card catalog in Wilson Library is a single dictionary arrangement file composed of main, added, series, and analytical cataloging entries. To derive a title sample profile directly from this catalog would require extensive random searching and selecting of only main entry cards (the only card assured of having complete data including tracings of all added entries, analytics and cross references). Therefore, we chose to derive a sample from the various shelf lists which together comprise the content of the union card catalog. We have estimated a gross total of 1.27 million titles are present in the shelf list. The sample later revealed title duplication so that the net total titles was adjusted to 1.25 million.

Sample Derivation

The Wilson shelf lists are divided according to the General Library and subject departmental libraries. A separate shelf list for the L.C. call no. libraries (Ames, Middle East, and East Asian) is maintained in a simple call no. order. 1270 drawers comprise the total, with 900 of these drawers representing the general library portion of the shelf list. The sample was selected as follows:

1. The total number of drawers was counted for each separate shelf list and 5% of the drawers were selected from each shelf list according to a random number table. With rounding or values representing less than 1 or a fraction of any number greater than 1 this resulted in 5.7% of the total drawers being used to select the random sample.

2. Once the drawers were selected, another order of single random digits was used to determine the point from which a group of cards in each drawer.

3. Then a group of 8 or 9 titles' cards were pulled from the selected point. If the point was greater than the available inches of cards in the drawer the number was divided by 2 unless it originally placed the sample point at the end of the available inches of cards in the drawer. In that case the cards were taken from the end of the drawer moving forward until 8 or 9 titles were produced.

4. A total sample size by number of titles was determined by estimating that the shelf lists together contained approximately 1,270,000 cards, the vast percentage of which were single shelf cards representing one title. Then it was determined that a manageable size sample would have to be employed due to the amount of effort which could be expended to perform the study. A 1% total sample would have represented 12,700 titles - far too many for one individual to use in data file comparisons in a manual mode within limited study costs. A .0005% sample results in 635 cards which is a manageable number with which to deal.

Testing Sample Validity

As the sample derived was random but very small we decided to compare known cataloging statistics for the past decade to the sample profile for L.C. copy, languages, and original cataloging. These comparisons show close correlation, with explainable differences for Spanish and Russian languages more prevalent in the later years. As a further check we performed the following experiment.

We compared our derived sample to the author/title portion of the University of California, Five Year Union Catalog Supplement 1963-1967 (U.C. catalog), extracting all samples from our sample with imprint dates later than 1967 (13.9%). This represented those titles not possible to find in the U.C. catalog as they were published after the inclusion period of this catalog. This lowered the total number in our sample to 543 possible matching titles. From this we found 31% of our sample matched the U.C. catalog.

As we found these matches we photocopied these pages - achieving a similar random sample of pages from the U.C. catalog to bring to Minnesota. From these random pages we selected several groups and compared all entries on those pages - omitting the entries that had been matched in the previous sampling process. Each group so matched produced from 31% to 32% overlap. From these two different comparisons using the same file and the close match of the characteristics of the derived sample to characteristics shown in the University's own cataloging statistics we conclude that the sample has sufficient validity on a practical basis for overlap comparisons for this study. Overlap figures derived using the sample will be conservative, particularly due to not attempting to reconcile entry variations during any given comparison. Our estimate, therefore in some cases, may be from 5% to 10% under what would be derived if every suspected variant cataloging entry would have been checked.

Sample Characteristics and derivation of the union catalog's characteristics

Table 6 gives characteristics of the 629 title sample shelf list cards together with their pertinent derived number of titles to which the characteristic applies. In some cases rounding yields more than 100%.

Table 6. Sample Characteristics

	No.	Percentage	No. of Titles in catalog
Cataloging Copy Source			
MARC	51	8%	100,000
L.C. card (Non-MARC)	398	63%	787,500
NUC	20	3.1%	38,750
Subtotal Supplied Copy	469	74.1%	926,250
Original cataloging	162	25.9%	323,750
Abbreviated shelf list cards	106	16.8%	210,000
Titles requiring more than 1 shelf list card (continuous cards)	16	2.5%	31,250
No. of shelf list cards in sample having subject & a.e. tracings	459	72.9%	911,250
Those with 1 subject heading	225	49.0%	446,513
" 2 "	120	26.3%	239,658
" 3 "	54	11.7%	106,616
" 4 "	9	2.0%	18,225
" 5+ (6 headings)	1	.2%	1,823
" 0 "	50	10.8%	98,415
Those with 1 added entry (non-subject)	278	60.5%	551,306
" 2 "	84	18.3%	166,759
" 3 "	21	4.5%	41,006
" 4 "	5	1.0%	9,113
" 5+ "	5	1.0%	9,113
" 0 added entry tracings	66	14.7%	133,953
Languages of catalog entries			
English	428	68.8%	860,000
German	53	8.4%	105,000
French	35	5.5%	68,750
Swedish	27	4.2%	52,500
Spanish	18	2.8%	35,000
Portuguese	19	3.0%	37,500
Danish	15	2.2%	27,500
Russian	11	1.7%	21,250
Chinese	8	1.2%	15,000
Italian	6	.95%	11,875
Norwegian	5	.79%	9,875
Arabic	2	.31%	9,875
Bulgarian	1	.15%	1,875
Yugoslavian	1	.15%	1,875
Dates of publication titles			
1968 - to date	88	13.9%	173,750
1960 - 1967	103	16.3%	203,750
1950 - 1959	113	17.9%	223,750
1940 - 1949	97	15.4%	192,500
1900 - 1939	172	27.3%	341,250
1800 - 1899	45	7.1%	88,750
Pre 1799 -	2	2.0%	25,000
Form of main entry heading			
Personal name	578	91.8%	1,147,500
Corporate name	30	4.7%	58,750
Conference or meeting	2	.5%	6,250
Title	19	3.0%	37,500

The incidence of title only added entries was also checked in the sample. 238 titles of 278 with a single non-subject added entry had a title added entry, i.e. 85% of single added entries will be title. Further, 85% of titles in the sample were found to have at least one added entry (non-subject) and 89% will have at least one subject heading. Only 3% of titles having 2 added entries (non-subject) have both versions of a title added entry.

We have excluded the theses shelf list from our Sample as this represents 100% original cataloging which could not be derived from pre-machine readable sources, as well as a special subset of the collection which may not be included in an initial system. Moreover, as the departmental shelf lists are not consolidated, each separate copy of a work has a card. Fortunately our random sample derived the percentage of overlap of duplicated titles - 2%. Therefore, we will use the adjusted estimate of 1.25 million titles as the population which would be affected by the sample data percentages.

Applying the data categories and percentages in Table 6 results in the title number totals in the No. of Titles column. Again rounding may produce a slightly greater total in number of titles than the estimated 1.25 million.

MARC II file overlap with the University of Minnesota Union Card Catalog

From our shelf list sample it was determined that 8% of the total estimated titles already cataloged would produce 100,000 records obtainable from MARC. This figure was determined from the 1968 imprint cutoff and L.C. cards identified as MARC. However, there are 17,000 records which the Library of Congress added to the MARC data base with Pre-1968 imprints. Therefore, although it cannot determine precisely how many of these "popular titles" are in the University's collections, it could be safe to assume that most of them would be present. Therefore, between 100,000 to 117,000 titles would be derived from MARC as of December 1974, including English and French languages. For this study the more conservative figure of 100,000 has been used.

For current cataloging purposes for the period 1975 and the future it has been determined that MARC record usage amount can be established by examining the cataloging statistics for the last available reporting year. This has been done assuming a constant number of acquisitions and a similar programmatic emphasis within the collections.

If it is assumed that 95% of the new titles cataloged are current imprints, or imprints not over a few years old the data in Table 7 results, using 1973/74 statistics. From this data a range of 66% to 81% of our cataloging records should be procured from MARC.

As a check on these annual statistics another source - a three month study of titles cataloged conducted by our Cataloging Division in 1974, was examined. This study revealed that during that period 85% of the titles had L.C. copy available.

If we note that the difference between foreign language cataloging and original cataloging is 8%, it may be assumed that this represents the amount of foreign cataloging within the supplied copy cataloging of 74%

which would produce 66% as a minimal percentage of English cataloging. Note that French, German, Spanish, Portuguese languages represent 15% of the total amount of cataloging and 43% of the total foreign languages cataloged. Adding this to the 66% base produces a probably expectation of 81% when MARC coverage extends in 1975 to these languages.

Applying the 95% rule would give us an adjusted total against which these percentages can be applied to derive a potential number of MARC II cataloging records used each year. For the period 1975 until further language coverage to Russian were accomplished, between 29,336 to 36,004 records would be obtained from MARC II machine readable files.

With the addition of Russian language in the period of 1976 or later, the percentage would rise to about 85% of current imprint cataloging, with lower retrospective imprint percentages in non-machine readable form for supplied sources until virtually all L.C. supplied cataloging records would be in machine readable state by 1980. With the broadened prospects of a shared cataloging input program for the future it would appear certain that 85% to 90% of our titles would be covered by MARC in the post 1976 period assuming present acquisitions patterns remain in force.

Table 7. 1973/74 Cataloging at the University of Minnesota

Total titles cataloged - 46,788

Adjusted current imprint total (95%) - 44,449

Catalog Copy Source	Percentage	Potential Total	95% (Adjusted total)
Library of Congress, N.U.C., NLM	74%	34,623	32,893
Original cataloging	26%	12,165	11,556
Total	100%	46,788	44,449
English language above	66%	30,880	29,336
Foreign language above	34%	15,908	15,113
French, German, Spanish, Portuguese (15% of total cataloging)	43% of foreign	7,018	6,667

Appendix 2. Detailed findings for MARC II monograph cataloging sources as found in Section 4.

Library of Congress MARC II Monograph Current Cataloging Data Sources

Library of Congress MARC II monograph cataloging records comprise the main source of current machine readable data available for a library to use in selection, ordering, and cataloging support. The MARC II monograph file contains 500,000 records as of December 1974. Except for approximately 17,000, all are records having imprint dates of 1968 to date. English language imprints alone were placed in this file until 1974 when French language were added. Beginning in January 1975 the Library of Congress has cataloged German, Spanish and Portuguese entries for inclusion in the file. The monograph file is available either for English language or All languages as a subscription service.

Procuring MARC II monograph cataloging information is complicated by the availability of a number of secondary sources as well as the subscription service itself.

In order to make a choice, alternatives must be costed to determine the actual total cost of acquisition and usage of the data for each given system. For the University of Minnesota system these alternatives and their costs have been examined here. Table 2 in Section 4 shows various component costs discussed below.

Weekly MARC II tape service subscription and compilation of a cumulative MARC II file.

If a library chooses to subscribe to the English or ALL LANGUAGE tape service, the records so obtained would need to be added to a cumulative MARC II file to permit searching over time in the system.

The University of Minnesota has acquired the source weekly English language tapes, with the initial two years (1968 and 1969) in quarterly 2400 foot reel form. The remainder are on 600 foot mini reels, i.e. 260 mini reels plus ten 2400 foot reels. The popular titles, RECON, and French language records have not been acquired. Because these tapes contain a variety of record types (new records, corrections to new records, deletions of new records, CIP records, corrections to CIP records, deletions of CIP records, CIP records updated to full MARC) it can be estimated that these records number 1 million. On the output side these records would produce 500,000 permanent MARC records comprising a cumulative MARC II file.

Therefore, the creation of a cumulative file from these weekly tapes, must be costed. Next, its maintenance on the on-line system, and its continued updating via either direct or indirect current source of MARC records must be costed. Then costs of alternate sources of already cumulative MARC II Source files or records can be compared, with the least costly procurement chosen.

Creation of Cumulative MARC II file from weekly tapes

The method proposed for costing the creation of this file assumes use of the University of Minnesota IBM 370/145VS machine at $200 per hour (clock time on/off).

The task is large enough to require separate time segments, where each may be restricted to no more than two hours. Therefore, a restart capability which does not require restoring data already processed, must be built in. Thus, the solution depends upon producing partially ordered data lists, i.e. a number of data lists where each list is ordered but no order exists across the lists.

The following method appears to offer the best solution for the present:

1. Read an input record.

2. Build a sort key comprised of
   a. L.C. card no. - 8 digits
   b. Track address - 15 bit binary number
   c. Type of record code - 1 character
   d. Data list number - 1 character

3. Search a cumulative ordered list of existing sort keys for the presence of the new key:
   a. If the key is found, change the type of record code; change the data list number; get the given track from the disk and replace the old record with the new. (keys for deleted records are deleted from the list).
   b. If the key is not found write the record to the end of the record file, store the track number in the sort key and insert the key into the existing list.

4. Test the threshold on the number of elements in the sort key array, and if not exceeded, return to Step 1. Otherwise,

5. Proceed successively down the sort key array, get the record at the given location and in this manner write all records, preceded by their sort keys, to tape.

These five steps generate one of the data lists. The total number of lists is therefore dependent upon the size of one list, which is the maximum number of 12 character elements possible to hold in computer core memory and process within the time slot available.

After the set of data lists is produced one pass has been made over the 270 tapes comprising the data base. These data lists must now be merged. It should be recognized that record redundancies still exist among the data lists even though internal to each list there are none. Since only seven tape units are available for input (one for output) we can only perform a seven way merge.

The logic of the merge phase requires a redundancy test in addition to the usual function of selecting the lowest key among the seven input buffers. Thus, for equal keys, the record containing the greatest data list number will be written to the output tape and the others ignored. The second pass over the data base will be completed when the last seven way merge is finished. Successive passes and merges are done the same way except for the last one when the sort keys are removed.

Only a rough approximation of processing times can be determined. The basic parameters are the number of records in the first pass and the size of the data list. It is assumed that the size of the data list determines the number of records possible to process per second (the time needed to build the list is large enough to exceed the tape time). Further it is assumed that the time to write a record to the disc file overlaps the time to read the next record from tape. A list of 16,384 elements (196,608 bytes) appears reasonable using this method and it is doubtful whether more than 8 - 10 records per second could be processed. Using these figures the first pass would produce:

$$\frac{1{,}000{,}000 \text{ records}}{16{,}384 \text{ records per data list}} = 61 \text{ data lists}$$

The actual number would be less than this after removing whatever redundancies existed within lists, but cannot be less than 31 if all redundancy were eliminated. If about 25% redundancy were eliminated then some 47 lists remain. Thus the production of each list would process about 22,000 records to yield the 16,384 records output. That would require about 45 minutes of IBM 370/145 time. With system overhead added it is conservative to assume over 60 hours would be needed for the initial pass since the handling of 270 tapes input alone is required.

The number of merge passes will be three since 61 divided by 7 = 9 data lists output from the initial merge pass; 9 divided by 7 = 2 data lists from the second pass, and 2 divided by 7 = 1 from the final pass.

Since each merge pass will be done close to tape speed the time required will be made up of rewind times and dismount and mount times as much as actual processing time. If 21 records per block (equal to one track) are assumed with an average of 636 characters per record given by L.C. then one tape can hold about:

$$\frac{2400' \times 12''}{13030 \text{ characters/block}}$$

1600 characters/inch + .5 = 3,340 blocks, or some 70,000 records

It will take some judicious planning to determine an algorithm for distributing these data lists among the tape units and physical tapes. Only 4 data lists maybe stored per tape, so that a minimum of 16 tapes would be produced from the first pass. As the final merge pass would produce 8 physical types for the 500,000 output records it is assumed that an average of 13 tapes will be read and 13 tapes will be written for each of the three merge passes or a total of 78 tapes will be handled. If it is assumed that ten minutes are required to read (write), rewind, mount (dismount) one tape then 13 hours would be needed for all the merge passes.

The costs of this are:

File processing 73 hours @$200 hr.	=	$14,600
Programming time 3 mo. @$1200 mo.	=	3,600
Test/debugging 4 hrs time @$200 hr.	=	800
Magnetic tapes 78 @$12.00 each	=	936
		$19,936

With University of Minnesota usable records estimated at 100,000 of the 500,000 produced we determine a per record cost to produce $.04 and a per usable record cost to produce $.20.

The annual cost of the MARC II tape for the 1975 volume year is estimated to be:

$1,500 for English language
$2,000 for ALL LANGUAGES

Updating a resident cumulative MARC II file with the weekly MARC II tape and residency cost of the file

Weekly Updating

This processing would take place on one minicomputer processor at one of the sites within the University of Minnesota Libraries system. An 800/1600 bpi 9 channel tape drive would be required for input depending upon the density chosen from the Library of Congress and other tape needs in the system. Since such a drive would be used for other purposes it would not be proper to cost the total value of this drive against this application. Rather, a total minicomputer system resource value per hour has been determined for our Bio-Medical Library system. This resource value assumes purchase of hardware, amortization of equipment over a life of five years and equipment replacement fund accrual. Maintenance charges also are included. This value is presently $28.00 per hour. Another system of this general type would have values of a similar nature.

Weekly processing requires reading an input tape of from 1500 - 2500 records on a 600 foot tape reel. Approximately two to three hours maximum would be required to process such a tape to input the records, locate appropriate records on the disc, insert, correct, or delete record and complete the process. The cost to do this would be 3 x $28.00 = $84.00 plus the programming costs to support this application. At most three man months of programmer time @$1200 per month for $3,600 would be involved. Therefore the total first year cost would be $4368 + $3600 or $7968 and the continuing costs $4368 per year. The initial year per MARC record costs and per usable MARC record costs would be $.13 and $.27 respectively. The continuing per record costs would be $.07 and $.15 respectively.

Residency Costs of a Full MARC II File

Currently the full MARC II monograph file of 500,000 records would require 350 million bytes of disc storage. Its annual growth rate is approximately 60,000 records or 2 million bytes of storage. We assume that 100,000 of these records up to December 1974 would be usable, i.e. overlap with titles presently in our collections. Based on the Bio-Medical Library PDP 11/40 system with 40 million byte disc drives nine drives and two controllers would be required at a cost of $203,600. However, based on use of Digital Equipment Corp. 88 million byte drives (RJP04), 4 units including one control unit would be required at a cost of approximately $113,000. Moreover, many other disk units are entering the market with similar or larger capacities and with 30% to 40% lower prices. Therefore, the Digital Equipment Corp. RJP04 units represent a maximum cost and best immediate alternative at this

writing. The RJP04 cost per MARC record stored would be $.226 or costed on the total usable $1.13 per record.

The growth rate of the MARC II file, assuming no truncations of the file, would fill an additional 88 million byte drive every two years with 8 years passing before an additional controller would be required. Therefore, approximately $13,000 annually in increased disc space would be required at a per MARC record cost of $.23 and per estimated record of the 29,336 annual usable records of $.44.

In the above file size estimates an L.C. card number ordered file with truncated search key access via author and title has been assumed which would require approximately 5 million bytes of storage.

Annual contract maintenance on the four 88 million byte disc drives and controller is estimated to be $800 per mo. or $9600 annually. This figures out to be $.02 for the complete MARC II file and $.096 per usable MARC record. Added incremental cost per year would be the maintenance of each added disc drive, estimated to be $2400 more for each of the first two years or $.04 per additional MARC record and $.08 per usable MARC record.

Hennepin County Library Cumulative MARC II file of English Language records

The second alternative to acquire a cumulative MARC II file would be to procure an already cumulated file. The Hennepin County Library and the University of Minnesota are cooperatively securing MARC II subscription tapes for the period April 1, 1975 - March 30, 1976 volume year. As part of its systems plans the Hennepin County Library has cumulated MARC records through use of the New York Public Library Catalog system software they are now installing. This tape is presently in the New York Public Library MARC II format. In this format an extra leader precedes each record and contains the L.C. Card Number, with the NYPL I.D. No. residing in TAG 001. Hennepin County is currently programming for the MARC communications format so that either format would be available to us through this cooperative arrangement. The cost of procuring this file will be the price of the annual MARC subscription - $2000 plus sufficient reels of tape (estimated previously at 8 reels using 800 bpi recording and blocked) or $96.

A programming estimate to process this tape file onto the PDP 11/40 system would be 3 man months @$1200 or $3600. System time for producing the file would be 10 hours at $28.00 per hour resource value or $280. The total cost to acquire and place this file on the PDP 11/40 system will be $5,976 exclusive of the disc storage costs.

Blackwell North America (formerly Richard Abel Co.)

Blackwell North America offers a variety of machine readable records. A full MARC II file is available along with two other files called MARC Sublevel 5 and MARC Sublevel 6. These files comprise an additional estimated 110,000 titles beyond the 500,000 L.C. MARC II titles. Moreover as of February 1975, Blackwell has acquired an additional file - the University of California Five Year Union Catalog supplement (covering 1963 - 1967) which contains approximately 750,000 bibliographic records. They plan to cleanse the California file and bring these records up to a higher standard of completeness. The following information was supplied by Blackwell or derived from sample tape dumps from each of these files. In general, this vendor offers

flexible price quotations tailored to each user's particular needs and procedures of identifying wanted records.

Sublevel 6 file Format

Both blocked and unblocked data can be supplied consistent with the format of the BLACKWELL L.C. MARC II file information. The customer many specify the length of block. These records are preceded by a four byte Leader in addition to the standard MARC II Leader. EBCDIC character set codes are also supplied for blocked records. Unblocked records can be provided either in ASCII or EBCDIC and do not have this four byte Blackwell Leader.

In these tapes the L.C. MARC Leader data has been altered slightly. The Encoding Level code (telling whether the record is an L.C. MARC II, or Blackwell Sublevel 5 or 6) has been shifted to byte 8 of the Leader. Byte 9 contains a modulus 11 check digit for the Blackwell Record I.D. Number. Bytes 17 - 23 store the Blackwell record number as an aid to searching. TAG 001 also contains the Blackwell Record I.D. number and its check digit with the L.C. card number moved to TAG 010.

Blackwell Sublevel 6 records contain the following TAG 008 fixed fields: Date entered, type of publication code, Date 1, Date 2, Language Code, Modified record indicator, Cataloging source code (also specifing if Blackwell modified L.C. cataloging or performed its own cataloging). All other L.C. MARC II fixed fields are not supplied. The variable fields supplied have the L.C. MARC indicator values and subfield structure changed in some cases. For example, subfields have been omitted, forcing the total information under a specific TAG to appear only under subfield a. Such is the case in TAG 260 Imprint, where place, publisher, and date are only separated by their punctuation. To break out such information into specific fields within the University of Minnesota system will necessitate character by character scanning of these records, performing computer aided editing and/or final editing of many records visually via CRT terminals. Although not explicity identified, all of the basic information required to print catalog cards and create search indexes appears to be present in the Sublevel 6 records.

Sublevel 5 file

The physical formats available on tape are identical to Sublevel 6 records. This file essentially represents titles from the National Agriculture Library (NAL) converted by the predecessor firm Richard Abel according to NAL's modification of the L.C. MARC II record. The MARC II Leader is modified in a similar fashion as on the Sublevel 6 file records. Also the Blackwell Record I.D. number and L.C. Card number are accomodated in an identical fashion. The same TAG 008 fixed fields are supplied as in Sublevel 6.

The change in this record is most apparent in the Variable Fields used. Not as extensive a set of L.C. MARC II TAGS are supplied. Indicator values are supplied for the TAG 245 Title, TAGS 4xx, TAG 505 Contents Note, and TAGS 6xx. These are simplified somewhat from the L.C. MARC II set. However, within the supplied variable fields subfield codes as used by L.C. MARC are maintained for virtually all of these fields as they appear on the file listing. In general Sublevel 5 records appear to be tagged more specifically and would require less computer or human editing to be converted into the University of Minnesota system.

Blackwell L.C. MARC II file

This file also has the four character Blackwell leader preceding the records and modification of the positioning of L.C. MARC II Leader fields to accommodate the Blackwell changes described under Sublevel 6 file. The L.C. card number has been moved to TAG 010 and this tag has been placed immediately after TAG 001, before the TAG 008 fixed field data.

There would be some minor programming required to change these records back into a true L.C. MARC II communications format for processing using such programs, or to convert this file directly into a University of Minnesota record structure. Programming to accommodate all three of these files would be very easy since their physical structure is virtually identical or can be determined by knowing which type of record is at hand for processing.

Overlap with University of Minnesota Shelf List Sample

To determine overlap this investigator provided a copy of the University's shelf list sample for Blackwell's comparison to their files. The MARC sublevel files and the L.C. MARC II file comprised the Blackwell data base at the time of comparison in January 1975. Blackwell found a 26.9% overlap which produces a total potential of 164,090 usable titles. As we have determined a minimum L.C. MARC II file overlap of 100,000 records it would appear that approximately 64,090 of these records would be produced from the Sublevel 5 and 6 records. These would require some degree of computer editing as well as human checking.

The price of these records would depend upon the volume to be purchased, with additional discounts of 3% for prepayment. In addition the complete Blackwell Data Base would be available for a purchase of $45,000. The maximum, undiscounted, low volume (6,000 titles) per record price for records found in the data base is $1.45. This price is for a Blackwell file conversion in which a library's shelf list cards are microfilmed and then brought to machine readable form by Blackwell's own conversion methods. It would appear that at the purchase price of $45,000 the 610,000 record size file would cost $.073 per record. Based on the usable estimate of 164,090 titles, the per record cost would be $.274 or close to the ceiling price of $.25 per record below. Blackwell has quoted $.25 as a ceiling price, for a retrieval of record by L.C. card No. from their files. Further price reductions to a $.10 per record level would occur on volume, whether MARC or non-MARC, and method of selection. Therefore, the cost per record would be between $.10 to $.274 depending upon pricing and selected method.

Similar costs to process these records would be incurred on the Minnesota system as those for handling our own cumulative MARC II file or the Hennepin County MARC II fiel. However, those records not L.C. MARC would require additional editing via computer and human methods at the time of their inclusion in the permanent on-line catalog in our system. These costs would be similar to those for in-house conversion operations.

Blackwell offers other services which could be attractive to a library doing retrospective conversion or current cataloging. One of these is the L.C. card number search which is priced at $.45 each if the record is found and $.02 each if no record is produced. Such a service on an annual recurring basis would cost the University of Minnesota approximately $17,000 annually based on our current cataloging volume and percentage of L.C. cataloging.

Information Dynamics Corp. (IDC) BIBNET file

The BIBNET machine readable data base is composed of the L.C. MARC II file and L.C. card number indices to other L.C. cataloging since 1953. We have not been able to ascertain with any degree of accuracy the number or nature of their other machine readable records. Moreover, the on-line service to access BIBNET offered via System Development Corp. consists of the L.C. MARC II file and some associated brief title index records. For potential use in the University of Minnesota system the BIBNET/LIBCON on-line file has been considered as a secondary L.C. MARC II source although there are apparently some other machine readable records available from IDC.

Although BIBNET makes provision for entry of a customer's record in the system from among those found in the BIBNET Title Index file, only a rough approximation of how many records have potential applicability to the University of Minnesota can be made. As 48.1% of our titles have imprint dates from 1950 - to date, with 75% of these L.C. derived, that would produce a maximum of 432,900 titles. At a conversion cost of $.90 each, a contracted conversion would entail $389,610 if the full number were produced.

The more intriging prospect for this type of on-line file is as a direct high speed transmission input source to the University Libraries own PDP 11/40 computer system. High speed transmission would be required, as at acoustic coupled data rates (300 Baud) the transmission of 100,000 MARC records would occur at the rate of approximately 155 per hour. Transmission of all applicable MARC II records would then necessitate 2,793 hours of connect time at $125,685 according to BIBNET/LIBCON 2701 system rates of $45.00 per connect hour. Service charges of $.90 per record would result in $90,000 in additional cost. Even the BIBNET 2707 service at $75.00 per hour would amount to a cost of $209,485 for 100,000 records, or only $6210 less than the 2701 service total. Higher speed transmission would lower these costs significantly. On this transmission basis and our present annual volume of needed records, an annual minimum of 190 connect hours would be used for future MARC transmission costing $8550 plus $26,402 in service use charge. Using the BIBNET 2707 service at $75.00 per hour the cost would be $14,250 and the better cost low speed alternative. This would be $.48 per MARC record transmitted.

Except for a more expensive receiving modem on our system, virtually the same communications hardware would be required for high speed transmission. This hardware consists of an auto-dial interface, auto-calling unit, line signal conditioner and line adapter, and would require one port of the sixteen available in one programmable multiplexer. Including the modem for higher speed transmission would involve a hardware investment of approximately $5,000 plus the leased telephone line charges. Recurring costs would be less than $1,000 maintenance plus telephone line charges.

Programming our PDP 11/40 system to accept direct digital transmission from the BIBNET system would involve simulating the calling protocol of the existing terminals such as the Data Point 2200 with cassette unit so that the BIBNET host computer would recognize our system as another of its terminals and send the data to core storage within our system. Then our programming would assume control and route the record to appropriate disc storage location for manipulation at our discretion. Six man months @$1500 per month would be required to design, program, test, and put into operation such a link.

On-line transmission of the 100,000 record portion of the MARC II file would be prohibitive at low speed rates and on the current service use charge basis for either the 2701 or 2707 service. However, high speed transmission findings are explained in the following portions of this section.

Further investigation of higher speed transmission rates was done to see if a more economic form of transmission could be used. A private telephone line from Minneapolis to Santa Monica, California would rent for $1350 per month. Without line conditioning up to 4800 Baud transmission rate is possible. At this speed approximately 1550 MARC records per hour could be transmitted. With C2 line conditioning up to 9600 or approximately 3050 records per hour could be transmitted. The C2 line conditioning would bring the monthly rental for a private line to approximately $1400 per month.

On the above basis at 4800 Baud about 65 hours of transmission time would be required for the 100,000 retrospective MARC II records. At 9600 Baud approximately 33 hours of transmission time for the 100,000 records would be required. As the private line can be used any time, the amount of time is equal to the number of days in each month or approximately 720 hours per month. On this basis we would have to transmit for at least one shift each day or about 240 hours per month for economic line utilization. The net cost of this unconditioned private line on an 8 hour per day utilization is about $5.60 per hour. About $364 of time would be needed at the 4800 Baud rate for the 100,000 records. On a C2 conditioned basis at 9600 Baud the $5.60 rate escalates to $5.83 with a drop in transmission cost to $192. But our real cost for such transmission is the monthly rentaltotal, with only a fraction of the time utilized. Even for the current project MARC II record usage of 30,000 titles per year this would require only about 2 hours of transmission time per month for the actual data. Probably an additional total of another 2 hours for searching of the other data base would bring the probably monthly need to 4 hours at 4800 Baud and just over half that at 9600 Baud. This would make monthly MARC II record transmission cost $.54 - $.56 per record, i.e. slightly higher than the 300 Baud service. To reduce this to an economic rate of $.10 per record would require 13,500 records minimum per month. This is already above the University of Minnesota volume by about 11,000 records per month. Therefore, private line transmission, for our volume, is too expensive according to the foregoing calculations.

Wide-Area Telephone Service (WATS) which would produce a 4800 Baud rate shows even higher costs. The one advantage here is the ability to call any U.S. city for the same price - $1635 per month for 255 hours usage. But with that rate the costs are also too high for the University of Minnesota volume. About 16,500 titles per month would give the necessary volume to bring down transmission costs to about $.10 per record.

Direct-dial long distance on a station to station basis enables transmission up to 4800 Baud with the user paying only for the actual time he requires. March 1, 1975 rate per minute to Santa Monica is $.42. The cost to transmit the 100,000 records would be $1,638 plus additional time for the searching process. If as much time were required to search as to send back, $3,276 would be the total cost or $.032 per MARC II record. In this mode it is assumed that a batch of L.C. card numbers would be built via terminal on the PDP 11/40 system disk file, then transmitted via a call to the BIBNET computer for processing. When ready BIBNET's computer would dial the University Library's computer which would prepare to receive the data for local manipulation.

The Library's annual needs for a 30,000 record total would require approximately 40 hours of time or $1,008 per year, including estimates for search time.

Therefore, on a transmission basis long distance dial up station-to-station batch request transmission would be most economic at 4800 Baud rate for the Library's volume. To this would have to be added the per record usage charge of $.90 which would bring the price per record from BIBNET to approximately $.94 each.

The equipment required to activate this connection to the Library's PDP 11/40 would involve approximately $5,000 including automatic calling unit, line adapters, and 4800 Baud modem. Approximately $200 per year hardware maintenance would be needed. Over an expected 5 year system life this amounts to an additional investment of $.04 per record or $1,200 per year. Adding this would bring the per record cost to $.98.

Ohio College Library System (OCLC) File

The OCLC system file represents over 1.2 million cataloging records. Of these 500,000 are L.C. MARC II records, the remaining portion being shared cataloging input from member libraries. Some of this shared cataloging is transcribed from L.C. and N.U.C. sources in exact replication, some with editing, and some records are original cataloging by the specific entering institution. The specific numbers in each category have not been determined for this study; however, each OCLC record gives the cataloging source and entering library. From the work in searching the University of Minnesota shelf list sample via OCLC terminal the records found appear to be typographically clean and a high number of full appearing catalog records, i.e. full collation statements, etc. Therefore, this file appears to represent a relatively high quality source of pre-machine readable records.

For the University of Minnesota the utility of this on-line file is most attractive as an on-line direct transmission source of MARC II records, as well as other contributed records. Therefore, for this study it is difficult to separate the OCLC file into MARC and non-MARC records other than for file size purposes.

Comparison of the OCLC and University of Minnesota Shelf List Sample

In comparing our sample we extracted the known L.C. MARC II records as these would be present on the file so that the file could be expected to produce 100,000 MARC records to December 1974 with an anticipated 29,336 minimum annual records beginning in 1975. Therefore, the problem became one of determining the overlap within the shared cataloging portion of 700,000 records. Our sample produced an overall 23.6% potential, or 165,200 records. Therefore a total of 265,200 records appear to reside in the OCLC file which the University of Minnesota holds.

As would be expected percentages of overlap vary with sequences of L.C. card numbers. Non-MARC L.C. card numbers in the 70-79 series produce a 28.5% overlap. L.C. card numbers in the 60-69 range produce 45% overlap, in the 50-59 range, 20% overlap and 1-49 range, a 17% overlap. Titles without L.C. card number, i.e. original cataloging or cataloging copied from an indeterminate source produce 20% overlap.

As previously stated the University of Minnesota would potentially use 48% of all MARC records planned for 1975 period or at minimum 29,336 records. This is 66% of our present cataloging load. In this analyst's opinion, with an anticipated 500,000 shared cataloging records expected from the OCLC system this next year, perhaps 100,000 records would represent 1970-to date non-MARC cataloging. If only 10% of this were usable by Minnesota, which appears a conservative assumption, then approximately 39,336 records minimum would be produced annually via the OCLC data base. This is 87% of our projected cataloging load based on our 1974 acquisitions. Therefore, the OCLC shared catalog portion not only offers significant potential for certain classes of retrospective records but can also be judged to significantly augment MARC II Library of Congress cataloging for the future - or until such time as MARC covers virtually all of L.C.'s Roman and Non-Roman alphabet cataloging.

OCLC File Quality

As previously described the sample derived from the shared cataloging input portions of the OCLC file appeared to be of excellent typographic quality, even for the one institution which several OCLC users we contacted said had poor quality records. Approximately 30% of these sample records did represent records where note information or other information fields were omitted or truncated. This was particularly disturbing when the original L.C. copy could have been input exactly as found on the original card. Most of this editing appears to be done to either conserve multiple card set use, number of entry points, i.e. card sets, or merely because the particular library used some short form cataloging standards such as abbreviated collation statements. If these records had been left in their original state and not so treated, the individual record quality of this file would be even higher - on a par with the New York Public Library Research Libraries Catalog. One caution - our valuation did not address entry authority forms or consistencies between families of records since the OCLC system does not contain built in authority controls. Therefore, such typical inconsistencies as appear in time in any library's catalog are bound to be present to some unidentified degree in this file as each institution creates records using their own authority standards to some degree.

OCLC Use in an On-Line System

Presently OCLC offers on-line terminal search and data entry capability and card production for its users. However, for those libraries using their own automat d internal procedures on a library dedicated computer system the main attraction of OCLC is as a raw source of bibliographic information for either current or retrospective cataloging work. Since such service is not currently offered by OCLC we can only give a best estimation of what such direct transmission service might cost as well as a comparison of it to the current service offered via OCLC 100 on-line terminals.

OCLC Retrospective File Costs

These costs to acquire tape data from the OCLC system via terminal searching have been based on costs attached to the OCLC service contract as of January 1975. If we assume searching of 265,200 titles in the OCLC file selected from our file as potential overlapping titles via publication date cutoff it would require 2 man years of terminal operation to retrieve and identify that number based on the current OCLC system response times.

Using the inclusive first time use charges of $2.11 for the estimated 265,200 records would cost $559,572 plus a batch tape copying charge of $265 and $80 in magnetic tapes including mailing. Even with an annual prepayment discount of 6% the figure would be $525,997. Therefore raw file cost would be $526,262.

Still to be added would be our in-house cost of editing and modifying these records on our own system.

OCLC On-Line Cataloging System Current Cataloging Costs

If the University were to use the present OCLC terminal and card production system we would use about 39,336 records initially and input only about 6,000 titles on which our reuse of the record would be free. Obviously our reuse of records would increase slightly based on our requiring corrected sets of cards or second sets of cards for subject library added copies. However, probably it would be about five years before any really significant shrinkage of first time use charges would occur. Therefore, on a 6% discount prepayment rate and inclusive first usage charge of $2.11 per record this system would cost Minnesota $77,885 plus $.034 per catalog card printed. Cataloging the total anticipated 45,000 titles and requiring an average of 12 cards per title would require producing 560,000 cards. This would cost $19,040, bringing an anticipated total OCLC annual cost to $96,925, or $2.15 per title cataloged.

OCLC Direct Computer to Computer Transmission Costs - Current Cataloging

The costs in this section have been worked out on a hypothetical basis since no such service presently exists. In this method the local University of Minnesota PDP 11/40 minicomputer system would be hooked via appropriate telephone line to the OCLC computer and would appear to that system as an OCLC 100 terminal. Terminals on the Minnesota system would access OCLC via our computer and upon return message transmission software would route the message to core memory for sending to the disc storage unit for processing locally. The programming of such a link requires the specifications for the OCLC communications protocols recognized by their computer and charcter set translation software on the Minnesota System and perhaps some appropriate local command modifications as well as conversion of the transmitted record to our internal processing structure.

The costs of such an approach would involve essentially the same factors as the similar hypothetical system prepared for BIBNET except that transmission line costs would be figured from Columbus, Ohio rather than Santa Monica, California to Minneapolis.

The rate for a private telephone line from Minneapolis to Columbus, Ohio is $650 per month. This would enable 4800 Baud service. On a C2 conditional basis this would rise to $700 per month for 9600 Baud service. On a 4800 Baud basis transmission of the expected total useful titles of 265,000 would require approximately 171 hours plus an additional amount estimated for search request transmission or 342 hours. With the OCLC hours of operation at 12 per day approximately 255 hours of service could be obtained in one month - so about 6 weeks of time would be used, i.e. $975 for 4800 Baud service or about .0036 per record. The same equipment at our computer would be required as for the BIBNET connection at a cost of $1200 per year adjusted on a five year system hardware life, i.e. $2175.

In addition to the above cost, under OCLC's present pricing schedules the charge for first usage of a record of $.904 would have to be added bringing the total cost to $241,735 or $.912 per record.

Use of C2 conditioned line would decrease the transmission time required for 265,000 records to 173 hours including the search transmission as well as the return data transmission. Three weeks of leased line time would equal about $525 worth used out of $700. On the $700 rate this lowers transmission to $.0026 per record. The total cost would be $241,460 or only a saving of $275.

Although these transmission rates are very low cost when compared to those for WATS it is necessary to examine the expected annual volume to determine which type of transmission method is really lowest cost for this volume.

Anticipating 39,336 records per year used from OCLC is roughly 3,278 per month producing a communications cost of between $.19 - $.21 per record. to this we must add the OCLC fee of $.904 per record and our $1200 worth of equipment for communications. This produces a total annual cost of $44,272 for 4800 Baud and $45,160 for 9600 private line service annually. Per record this is $1.125 and 1.148 respectively.

Therefore, the private conditioned line is not recommended as it is more expensive per record for our current volume.

The final alternative would be long distance dial up station-to-station at $.32 per minute as of March 1, 1975. Again speeds up to 4800 Baud can be achieved. The University of Minnesota's annual needs would be approximately 50 hours or $960 per year. Add to this the communications equipment at $1200 and the OCLC charge of $.904 per record produces an annual cost of $37,720 or $.958 per record. Clearly on this volume long distance station-to-station transmission at 4800 Baud is the cheapest alternative.

Since the kind of service we are describing here and costing is not currently offered we cannot include any costs to OCLC for developing the capability to handle non-private line 4800 Baud transmission, handle block transmission of a number of L.C. Card No. or search key searches together with the block transmission of the resulting records or no match messages. However, it would appear that this capability would be a desirable addition to make as such services would appear to have the greatest attraction to those users with their own computer facilities and not requiring the card production support services. This investigator urges OCLC to give serious consideration to implementing such service on a cost basis attractive to those libraries having their own in-house on-line computer systems.

Special Subject Files

Monograph cataloging information appears in some instances as part of certain subject periodical literature oriented data bases available for on-line search. In most cases these files would not produce sufficient titles to be considered viable sources of cataloging information. Moreover, information content or data element identification may be lacking in such files. The file described below appears to offer some useful cataloging information for medical libraries using MeSH subject headings and NLM classification.

The National Library of Medicine on-line CATLINE file is a special file of the bibliographic records contained in the Current Catalog of the National Library of Medicine from 1965 to date. This file contains data needed for monograph cataloging, but format and rules vary from that of the Library of Congress MARC file. As of December 1974 approximately 110,000 titles are available. Those titles in English and French are also covered by the MARC tape currently. Therefore, for a medical library using NLM classification and subject headings this is a good source to consider for those titles not covered under MARC currently.

Otherwise, CATLINE can be considered a source for medical monograph cataloging done between 1965-1967 for all languages and for other than English and French presently. With 1975, German, Spanish, and Portuguese are being added to MARC, further restricting the unique titles to be obtained in the future from CATLINE. It would appear to be a file most useful to pick up NLM specific information for use in medical libraries following their classification and subject headings as well as their specific form of entry variations.

Therefore, the use of this file will have to be made -- not on a cost basis but rather on whether NLM classification and subject headings are desired in the respective library's catalog access. If they are, then, dependent upon volume, either manual use of Current Catalog, on-line search via CATLINE, or automatic computer to computer handling of this file may be considered. If other on-line files are to be accessed via a minicomputer system--such as MEDLARS--then programming to handle this file could be accomplished as part of that source file handling, with provisions for writing out on our local file any desired retrieved CATLINE records for editing by our own cataloging staff.

Approximately $3,000 would be involved to program access to the MEDLARS file via our PDP 11/40 system so that it would appear that a cataloging volume of only 2,000 titles would make this investment as a by-product pay off after one year. This is based on an average cost of $1.30 manually to search/verify a title.

Otherwise, if CATLINE were not to be accessed as a by-product of other NLM access, this low volume added to the need for capturing only the NLM specific fields, i.e. classification and subject heads, or occasional entry variations and few non-MARC records, the cost would be higher than either a single present CATLINE search or Current Catalog search.

Appendix 3. Detailed findings for retrospective existing machine readable cataloging files as found in Section 5.

Decision Making Factors

How can a library decide whether use of another library's machine readable file will be an economic aid to its own conversion? The answer will depend upon:

How many usable records can be obtained from the file (using a conservative estimate),

The quality of the records,

The cost of acquiring the file,

The cost of programming to convert it, build search indexes, and do any reformatting required.

The cost of temporarily storing it for the duration of the in-house conversion process.

The costs established for the above must be compared to the cost of conversion of the library's catalog by some means that does not use a pre-machine readable file (other than MARC) and also determine the amount which could be saved through the use of another Library's file. Then the costs for these methods can be compared to determine which is lower.

It has been assumed that any retrospective files to be seriously considered for any University of Minnesota conversion effort would be those constructed for a research library. Public library files and undergraduate academic libraries have sufficiently different collections to make their catalogs quite different. As a simple test of this a comparison of the Hennepin County Library's computer produced book catalog with the University of Minnesota catalog was done and a 5% overlap was found. This would produce a maximum of 5,100 titles from the Hennepin County file. The cost to determine the applicable titles, programming, etc. would be more costly than the original entry of this number of titles. Therefore, only New York Public Library Research Libraries; University of California, Berkeley, Five Year Union Catalog Supplement, and the University of Chicago Library files were examined separately as potential retrospective sources. In Table 3 in Section 5 the OCLC shared cataloging file has been used for cost comparisons. Appendix 2 contains all OCLC related detailed information as it is difficult to separate their file into separate entities. Specific data on each file considered follows.

New York Public Library. Research Libraries Catalog File

As of January 1975 this file contained 204,317 titles. The record structure and data element identification is MARC II. The NYPL MARC tape format is identical to the MARC II communications tape except for:

1. A pre-leader portion containing the L.C. card no. in a fixed number of bytes.

2. Removal of the L.C. card no. from TAG 001 as above and substitution of NYPL record I.D. number. The L.C. card no. has been moved to TAG 010 as provided in MARC II.

Paul J. Fasana supplied four volumes of the Research Libraries Catalog to this investigator for use in the study. Random pages were chosen for comparison to the Minnesota union card catalog instead of our statistical sample as we did not have the complete New York catalog.

This catalog includes all cataloging performed from January 1, 1971 to date. It includes MARC II records, L.C. derived cataloging, and original entry cataloging. The quality of the records appears very high--probably a good deal higher than the University of Minnesota union catalog due to strict L.C. subject heading usage and fully automatic computer aided authority controls on the file. Figure 1. shows a sample page with a corresponding card from the Minnesota catalog.

Figure 1. New York Public Library Catalog Sample

Van der Post, Laurens.

A[illegible] the editors of [illegible] Books. Photographed by Brian Seed and Richard Jeffery. New York, Time-Life Books [1970] 208 p. col. illus., col. maps, ports. (part [illegible]) 26 cm. [illegible] of the world) [illegible] LC [illegible]

[Sc [illegible]]

A story like the wind. London, Hogarth Press, 1972. xi, 370 p. 22 cm. NN 73-[illegible]

[JFD 73-[illegible]]

Van Der Slik, Jack R., 1936-

(comp.) Black conflict with white America, a reader in social and political analysis. Edited by Jack R. Van der Slik. Columbus, Ohio, Merrill [1970] xii, 344 p. 23 cm. (Merrill political science series) Includes bibliographical references. CONTENTS. - The social and economic status of the Negro in the United States, by St. C. Drake. - Self concept, level of aspiration, and academic achievement, by M. D. Caplin. - Cops in the ghetto: a problem of the police system, by B. Levy. - Political attitudes of urban Blacks and Whites: some implications for policy makers, by E. F. Cataldo, R. M. Johnson, and L. A. Kellstedt. - The Negro voter: [illegible], by C. Stone. - Negro political strategy: coalition or independent power politics, by H. Holloway. - Black elected officials in the [illegible] exploratory analysis, by J. R. Feagin. - [illegible] the Federal employment opportunity and [illegible] S. Krislov. - [illegible] Porter and R. L. Crain. - Protest [illegible] by M. Lipsky. - The cities and Negro student protest, by A. M. Orum [illegible]

[IE 71-512]

(joint author) Force, George T. Theory and research in the study of political leadership. [Carbondale] 1969. xii, 90 p. NN 72-[illegible]

[JLF 72-[illegible]]

(joint author) Force, George T. Theory and research in the study of political leadership. [Carbondale] 1972. viii, 139 p. NN 73-[illegible]

[JLF 73-[illegible]]

Van der Spoel, [illegible] see [illegible]

Van der Spuy, Kenneth Reid, 1894- Old Nectar and [illegible]. Cape Town, Books of Africa, 1963. 170 p. illus. (part col.) 26 cm. NN [illegible]

Van der Spuy, Una. South African shrubs and trees for the garden. [illegible] ed. Johannesburg, [illegible]

Van der Leeden, Frits. Ground water: a selected bibliography. Compiled and edited by Frits van der Leeden. Port Washington, N.Y., Water Information Center [1971] [illegible] p. 24 cm. NN [illegible] LC [illegible]

[JSS 72-23]

This document was processed for the ERIC Document Reproduction Service by the ERIC Clearinghouse at Stanford. We are aware that some pages probably will not be readable in microfiche or in Hardcopy form. However, this is the best available copy, and we feel that the document should not be withheld from interested readers on the basis of these unreadable pages alone.

Z7016.55147 V224

Van der Leeden, Frits.

Ground water: a selected bibliography. Compiled and edited by Frits van der Leeden. Port Washington, N.Y., Water Information Center [1971]

viii, 123 p. 24 cm.

2. Groundwater flow. Bibliography. 1. [illegible] I. Title.

Z7016.55147 016.55149 73-[illegible]

45

Usable Records and Quality

Comparison between the NYPL sample and the University of Minnesota catalog showed that 31% of the entries checked in the Research Libraries Catalog were found. This 31% was composed of 49% L.C. MARC II titles, 44% L.C. non-MARC titles, and 7% original cataloging. This comparison did not attempt to check alternate forms of entry where the potential for such differences occurred. It is felt that such checking would not produce over 10% additional usable titles. Since a conservative estimate gives assurance that if use is cost justified on this amount any additional titles used merely would make the use of the full file more justified.

These percentages above produce the following usable records for the NYPL file:

Total 31%	63,338 records
49% L.C. MARC II	31,036 records of above
44% of L.C. Non-MARC	27,869 records of above
7% original	4,433 records of above

Since it is assumed that the Minnesota system would have an existing MARC II file, a potential of 32,302 new records would result through NYPL file usage. But if there were no existing MARC II file, only 31% of our retrospective MARC II file needs could be met via this file.

Acquisition Costs

New York Public Library has not quoted the terms of their file availability so it is difficult to assess the cost of acquisition. Obviously the physical costs of file duplication, and documentation, as well as a cost of the hardcopy catalog for ease of record usage would be minimal. Or, the maximal figure could be commercial costs.

We have chosen here to cost the file on a basis comparible to the costs published for the University of California Five Year Union Catalog data base, i.e. approximately .106 per record or for NYPL $21,657. As the acquisition cost is in doubt the other cost factors can be compared for these files and then these costs compared to those for in-house conversion as has been done in the Section 5. conclusions.

Programming Costs

It has been assumed that these costs would include the building of L.C. Card No. indexes, truncated author/title indices and converting the NYPL format to our own Minnesota internal format. Then the file would have to be stored on our system for the estimated 30 months our conversion effort was being carried out. An estimate has been made that 4 man months at $1,600 per month or $6,400 would be needed to program and create such a file. About $700 worth of system utilization time would be required so that $7,100 total would be required.

Temporary Storage Costs

The initial file would require approximately 168 million bytes of disc storage so two 86 magabyte RJP04 units would be required. This would equal $53,250 worth of purchased equipment which later could be put to other use. If the estimated life of such hardware is 60 months and equipment is to be

amortized over this period, a conversion effort of 30 months will mean that $26,625 or 50% of this total must be costed out to this purpose. This does not include $.03 per record maintenance for the last 18 months of the period. The full costs of using the NYPL Research Libraries file is shown in Table 8 with and without the maximum acquisitions costs considered.

Table 8. Costs of the NYPL Research Libraries file for use on the University of Minnesota System for Retrospective Conversion.

Usable Records total	63,338	
Usable Non-MARC II	32,302	
Usable MARC II	31,036	
	Acquisition costs	$21,657
	Programming costs	$ 7,100
	Temporary storage costs	$26,625
	Total costs	$55,382
Costs per non-MARC usable record		$ 1.71
Costs per total usable records		$ 1.05
Costs exclusive of acquisitions		$32,725 total
Costs per non-MARC usable record exclusive of acquisitions costs		$ 1.04
Costs per total usable records exclusive of acquisitions costs		$ .53

University of California, Berkeley. Five Year Union Catalog Supplement Data Base.

This file contains 750,000 estimated records with 350,000 having L.C. Card Numbers. The cataloging period covered is 1963-1967 or pre-MARC (except for the estimated 17,000 titles in the popular titles conversion). For our purposed here the file is virtually non-duplicative of MARC.

The format of this file is that of MARC but the tagging and identification of specific data elements varies. For example the L.C. Card Numbers are embedded in a 900 field along with other data elements and without subfield identification. The Institute of Library Research is planning to modify this condition via programming. Four our purposes at this time it must be assumed that a more complex programming effort would be required than for the NYPL file.

Usable Records and Quality

Extracting the MARC II samples from the shelf list sample we found a 31% overlap of possible titles. Using the cross comparison mentioned previously in Appendix 1. and as in the NYPL comparison a 43% overlap was found, suggesting on a very practical basis the goodness of our shelf list sample and general comparison technique. Using the 31% figure produces a potential 232,500 titles--all non-MARC for practical purposes.

The quality of this file though is poor when compared to the other files or the University of Minnesota catalog. Many keyboarding errors exist in significant fields, some records appear truncated, the number of duplicate records is not known but appears to be significant when scanning any given page. Figure 2 is a sample page from this catalog. In our opinion every record from this file would have to be closely checked before being used. Probably extensive hand correction to identify specific fields would be needed. Obviously, when one considered the methods used to compile this catalog we can only admire the degree of quality that was achieved, as once keyboarding was done all other processing was via computer program.

Examples of the kinds of errors seen on virtually any random page are:

1. Duplicate records for titles not found via program editing due to keyboarding errors, variant copy, omissions of characters in foreign alphabets and variant transliterations, capitalization of words and lower case letters both,

2. Collation variations in otherwise identical records,

3. Subject heading variations in otherwise identical records,

4. Omission of cataloging data fields which appear to be L.C. originated with full L.C. record following,

5. Use of a plural or singular form in otherwise identical records,

6. Omission of spaces between words in an otherwise correctly appearing record,

7. Records identical except for extra added entry made by one library or omission of series entry by one library,

8. Duplicate forms author with birth and death dates and without or with death date supplied,

9. Bracketed information such as place of publication and unbracketed on the same inprint and edition record,

10. Use of author statement following title and its omission,

11. A contents note and none on otherwise identical records,

12. Single insignificant keyboarding error in a long extract note in otherwise identical records,

13. Omission of letters such as CAIFORNIA or CA'IFORNIA for CALIFORNIA, and

14. Author date subfield code missing so dates appear as if part of middle name or initial.

CAIANIELLO, Eduardo R ed.
—Automata theory. Edited by E.R. Caianiello. New York, Academic Press 1966. 342 illus English or French. Includes bibliographies. 1. Automata—Addresses, essays, lectures. I. International Spring School of Physics. 6th, Ravello, Italy, 1964. 0010876. Q300 I54 1964—SC.

—Cybernetics of neural processes, course held at the Internation School of Physics, sponsored by NATO at the Istituto di Fisica Teorica. Università di Napoli. April 26-May 13, 1962. Roma. Consiglio Nazionale delle Ricerche. 1965. 395 p illus. (La RICERCA scientifica. Quaderni, 31) 1. Cybernetics. 2. Nervous system—Physiology. Biomedical Library 1022583. Q 310 C121c 1965—L.

—xiv. 342 p. 2400. *See* Automata theory. Edited by E. R. Caianiello. New York, Academic Press, 1966. Q 300 A939—SD S & E.

CAICARAS de Juquef. 1961. *See* Dad. Penf. PQ 9597 D25403—L.

CAICEDO Castilla, José Joaquín.
—El panamericanismo. Buenos Aires, R. Depalma, 1961. 484 p. 24 cm. 1. Pan-Americanism. 2. Pan—American treaties and conventions. 3. Law—America. 4. Law—Codification. 61-39962 1526464. F 1418 C23—SB.

—The work of the Inter-American Juridical Committee. Washington, Pan American Union, 1964. iii. 108 p. 28 cm. "This document is published by the Codification Division, Department of Legal Affairs." 1. Inter-American Council of Jurists. Inter-American Juridical Committee. 2. International law. 3. Pan—Americanism. I. Pan American Union. Codification Division. P A 65-6 1296182. Law Library—L.

CAICEDO, Daniel.
—Viento seco. 2. ed. Bogotá, 1954. 182 p. 18cm. 1319803. PQ 8179 C135 V5 1954—IR.

—Viento seco; novela. Con un prólogo de Antoinio García. 3. ed. Buenos Aires, Editorial Nuestra América, 1954. 184 p. 20 cm. PQ8179 C135V5 1954 54-44599 1403607. PQ8179 C135V5 1954—R

CAICEDO, Dorothy.
—These are the lists of ny despairs. New York, Philosophical Library [1966] viii, 87 p. 22 cm. Poems PS3553 A37T5 811 54 65-24272 1538507. PS 3553 A37 T5—SB.

CAICEDO Montúa, Francisco A.
—Banzay; diario en las trincheras coreanas. [n.p., 1961] 224 p. illus. 1. Korean War, 1950-1953—Personal narratives—Columbian. 1122123. DS 921 6 C125—L.

CAIDEN, Gerald E.
—Career service; an introduction to the hisotry of personnel administration in the Commonwealth Public Service of Australia. 1901-1961 [by] Gerald E. Caiden. [Carlton, Australia] Melbourne University Press; London, New York, Cambridge University Press [1965] xviii, 530 p. illus., ports. 24 cm. Bibliography: p. 508-520. 1. Civil service-Australia-Hist. JQ4047 C3 354 9400! 65-21860 1673045. JQ4047 C3—D

—Career service; an introduction to the history of personnel administration in the Commonwealth Public Service of Australia, 1901-1961 [by] Gerald E. Caiden. [Carlton, Australia] Melbourne University Press; London, New York, Cambridge University Press [1965] xviii, 530 p. illus., ports. 24 cm. Bibliography: p. 508-520. 1. Civil service—History. JQ4047 C3 354 9400! 65-21860 1225690. JQ 4047 C122c—L; JQ4047 C24—B JQ 4047 C3—SB

—The Commonwealth Bureaucracy [by] Gerald E. Caiden. [Melbourne] Melbourne University Press; New York, Cambridge University Press [1967] xiii 445 p. illus. Bibliographical footnotes. 1. Australia—Politics and government. 2. Civil service—Australia. 3. Bureaucracy. 2160997. JQ4024 C22—B

—The Commonwealth bureaucracy. by Gerald E. Caiden. Melbourne. Melbourne University Press 1967$London. New York.$Cambridge University Press$1967. 445 xiii, 445p. tables. Bibliographical footnotes. 1. civil service —australia. 0066577. JQ4047 C3—SC

CAIDIN, Grace, joint author.
—Aviation and space medicine. 1962. *See* Caidin, Martin, 1927- RC1062 C3—B Biology Library

CAIDIN, Martin, 1927-
—Air force; a pictorial history of American airpower, by Martin Caidin, in cooperation with the U. S. Air Force. New York, Bramhall House [c1957] 232 p. illus. 32 cm. 1. Aeronautics, Military—United States. 2. Air power—United States. 3. U. S. Air Force—Pictorial works. 1788963. UG 633 C133—SD

—Air force, a pictorial history of American airpower, by ... in cooperation with the U. S. Air Force. New York, Rinehart 1957. 232 illus. ports. 1. Aeronautics, Military-U. S. 2. Air power-U. S. 3. U. S. Air Force-Pictorial works. 1958164. UG633 C14—SC

CAIDIN, Martin, 1927-
—The astronauts, the story of Project Mercury, America'sman-in-space program. Drawings by Fred L. Wolff. New enl. i. e. 2d ed. covering the flights of Alan B. Shepard and Yuri Gagarin. New York, Dutton 1961. 224 illus. 1. Project Mercury. 2. Shepard, Alan Bartlett, 1923- 3. Gagarin, Iurii Alekseevich, 1934- 0057744. TL789 8 U5M44 1961—SC.

—Aviation and space medicine; man conquers the vertical frontier, by Martin and Grace Caidin. [1st ed.] New York, Dutton, 1962. 215 p. illus. 21 cm. 1. Aviation medicine. 2. Space medicine. I. Caidin, Grace, joint author. RC1062 C3 616 98021 62-7335 2113892. RC1062 C3—B Biology Library.

—Cross-country flying. Photos. by James Yarnell. Drawings by Fred L. Wolff. New York, Dutton 1961. 253 illus. 1. Air pilots-Correspondence, reminiscences, etc. 2. Aeroplanes, Private. 3. U. S.-Descr. and trav.-Aerial. 0057724. TL721 4 C29—SC

—Flying; principles of flight and the development of aircraft. New York, Holt, Rinehart and Winston [c1963] 144 p. illus. 21 cm. Bibliography: p. 139. 1. Aeronautics—Popular works. TL546 7 C2 629 132 63-22248 2110663. TL545 C19—B Educ.-Psych. Library.

—Hydrospace. Drawings by Fred L. Wolff. [1st ed.] New York, Dutton, 1964. 320 p. illus., charts, ports. 22 cm. 1. Oceanographic research. 2. Marine resources. GC57 C3 551.46072 64-21962 1539648. GC 57 C3—S3; GC57 C3—D

—Hydrospace. Drawings by Fred L. Wolff. 1st ed. New York, Dutton 1964. 320 illus. 1. Oceanographic research. 2. Marine resources. 0035191. GC57 C3—SC.

—The long arm of America, the story of the amazing Hercules air assault transport and our revolutionary global strike forces. Drawings by Fred L. Wolff. [1st ed.] New York, Dutton, 1963. xii, 369 p. illus., ports. 22 cm. 1. Hercules (Turboprop transports). UG633 C147 358 44 63-15777 1542019. UG 633 C3—SB.

—The man-in-space dictionary, a modern glossary. [1st ed.] New York, Dutton, 1963. 224 p. illus. 25 cm. 1. Astronautics—Dictionaries. TL788 C3 629.403 63-14274 1422095. P TL788 C3—R; TL789 C3 R—D; TL 788 C12m—L Engineering & Mathematical Sciences Library; Ref TL 788 C3—SB; TL 788 C146—SD S S E.

—The man-in-space dictionary, a modern glossary. 1st ed. New York, Dutton 1963. 224 illus. 1. Astronautics-Dictionaries. 0057735. TL788.C3—SC.

—Marooned, a novel. [1st. ed.] New York, Dutton, 1964. 378 p. 22 cm. PZ4 C133Mar 64-11082 2181425. 971 C133 mar—B; PS3553 A5 M3—D
—The mission, by MartinCaidin and Edward Hymoff. [1st ed.] Philadelphia, Lippincott [1964] 208 p. illus., map, ports. 22 cm. Bibliographical references included in "Notes and sources" (p. 197-203) 1. Johnson, Lyndon Baines, Pres. U. S., 1908- 2. World War, 1939-1945—Oceanica. I. Hymoff, Edward, joint author. E847 C3 940.5426 64-15630 2110771. E847 C25—B
—The moon: new world for men. Indianapolis, Bobbs-Merrill z[1963] 405 p. illus 24 cm. 1. Space flight to the moon. 2. Lunar probes. TL799 M6C295 629.4353 63-14583 2176919. TL799 M6C295—B; TL 799 M6C12—L Engineering & Mathematical Sciences Library.

—Overture to space. [1st ed.] New York, Duell, Sloan and Pearce [1963] 300 p. 21 cm. 1. Astronautics—Popular works. TL793 C29 629.4 63-10347 2182293. TL793 C295—B

—The ragged, rugged warriors. Drawings by Fred L. Wolff. [1st ed.] New York, Dutton, 1966. 384 p. illus., maps, ports. 22 cm. Bibliographical references included in "Notes and sources" (p. 371-375) 1. World war, 1939-1945—Aerial operations. 2. World war, 1939-1945—Pacific Ocean. D767 C23 940.5442 66-10663 1676359. D767 C23—D; D 767 C3—SB.

—Red star in space. [New York] Crowell-Collier Press [1963] 240 p. illus. 22 cm. 1. Astronautics-Russia. TL789 8 R9C3 629 40947 63-10949 2091998. TL789 8 R9C3—B

—Rendezvous in space; the story of Projects Mercury, Gemini, Dyna-Soar and Apollo. Drawings by Fred L. Wolff. [1st ed.] New York, Dutton, 1962. 320 p. illus. 21 cm. 1. Space flight. TL793 C3 629.45 62-14725 1122126. TL 793 C12r—L Engineering & Mathematical Sciences Library. TL 789 8 C133—SD

—Rendezvous in space; the story of Projects Mercury, Gemini, Dyna-Soar and Apollo. Drawings by Fred L. Wolff. 1st ed. New York, Dutton 1962. 320 illus. 1. Astronautics. 0057753. TL793 C3—SC

—Thunderbirds. Line drawings by Fred L. Wolff. New York, Dutton [1964] 256 p. illus. 1. United States. Air Force 4520th Aerial Demonstration Team. I. Wolff, Fred L. illus. LCR-mdg U C R 1439309. C133t—R JUV

—Vanguard. The story of the first man-made satellite. 1st ed. New York, Dutton 1957. 288 illus. 1. Project Vanguard. 0089887. TL796 C2—SC.
—Why space? And how it serves you in your daily life. Foreword by Homer E. Newell. New York, J. Messner 1965. 20? illus. 1. Astronautics and civilization. 1973438. CB440 C3—SC

CAIDIN, Martin, 1927-
—Wings into space; the history and future of winged space flight. New York, Holt, Rinehart and Winston [1964] 143 p. illus., ports. 21 cm. (Holt Library of science. Ser. 2, v. 24) Bibliography: p. 131. 1. Winged spacecraft. TL793 C33 629.45 64-23010 2074031. TL793 C33—B Educ.-Psych. Library

CAIDIN, Martin, 1927-The silken angels; a history of parachuting. [1st ed.] Philadelphia, Lippincott, 1964. 264 p. illus., ports. 24 cm. 1. Parachuting—History. TL750 C3 629.134386 63-20384 2114662. TL750 C23—B Educ.-Psych. Library.

CAIFORNIA cases on secured transactions relating to land. 1954. *See* Riesenfeld, Stefan Albrecht, 1908- ed. F852 .1 R55—B BANCROFT LIBRARY.

CAIFORNIA. (State) Citizens' Advisory Committee to the Attorney General on Crime Prevention.
—California jails; report to Attorney General Edmund G. Brown. [Sacramento] 1956. 70 p. 23 cm. 1. Prisons—California. HV8324 A6 1956 365.33 56-63544 2037575. HV8324 A6 1956a—B Documents Dept.

CA-IFORNIA. UNIVERSITY.
—Biology in the university, problems and prospects. 1963. *See* UNIVERSITY OF CALIFORNIA SPECIAL ALL UNIVERSITY CONFERENCE ON BIOLOGY, DAVIS, 1963. QH 315 U585 1963—L; QH315 U58 1963—D.

CAIGER, George.
—Tojo say no; Japanese ideas and ideals, by George Caiger. Sydney, London, Angus and Robertson ltd 1943. viiip., 3 ., 165, 1p. plates. "Typical specimens of the public story-tellers' art" collected by Caiger, translated from the Japanese under his direction, and the English versions revised and rewritten byhim. cf. Introd. Author's presentation copy. 1. Tales, Japanese. 2. Japanese literature—Translations into English. 3. English literature—Translations from Japanese. 4. national characteristics, japanese. 0087454. PL890.A2C3—SC.

CAIGER, George ed.
—The Australian way of life. Edited by ... under the auspices of the Australian Institute of International Affairs. New York, Columbia University Press 1953. 158 plates, tables. (Way of life series. Bibliographical footnotes.) 1. Australia—Social conditions. I. Australian Institute of International Affairs. S. Way of life series. 0031811. DU107.C3—SC.

CAIGER-SMITH, A.
—English medieval mural paintings. Oxford, Clarendon Press, 1963. 190 p. plates (1 col.) 25 cm. Bibliography: p. [183]-184. 1. Mural painting and decoration, English. 2. Mural painting and decoration, Medieval. ND2728 C35 751 730942 63-5641 1122127. ND 2728 C12e—L Art Library; ND2728 C35—D; ND 2728 C3—SB.

—English medieval mural paintings. Oxford, Clarendon Press 1963. 190 plates (1 col.) Bibliography: p. 183-184. 1. Mural painting and decoration, English. 2. Mural painting and decoration, Medieval. 0043430. ND2728 C35—SC.

CAIGER, Stephen Langrish.
—Bible and spade; an introduction to Biblical archaeology, by Stephen L. Caiger, B. D. London, Oxford university press, H. Milford, 1936. xii, 218 p., 1 l. incl. illus., tables. front., plates, facsim. 19 cm. Plan and map on lining-papers. "List of books" p. [204]-211. 1. Bible. O. T.—Antiquities. 2. Bible—Evidences, authority, etc. 3. Excavations (Archaeology). BS620 C3 221 93 36-27088 1677358. BS620 C3—D

—British Honduras, past and present. London, G. Allen & Unwin [1951] 240 p. illus. 22 cm. 1. British Honduras—History. F1446 C15 972.82 52-1545 1541266. F 1446 C25—SB

CAIL, Odile.
—La Chine. 187 reproductions en couleurs et en noir dont 145 photos. originales de l'auteur. Paris, F. Nathan [1967] 160 p. illus., map, 4 col. plates, table (Pays et cités d'art) 1. Art—China. 2. China—Description and travel—1949- 2143390. N7340 C2—B.

CAILHAVA d'Estandoux, Jean François.
—Études sur Moliere, ou Observations sur la vie, les moeurs, les ouvrages de cet auteur, et sur la maniere de jouer ses pieces, pour faire suite aux diverses editions des oeuvres de Moliere. Paris, Debray 1802. 355. 1. moliere, jean baptiste poquelin, 1622-1673. 0079593. PQ1852 C3—SC

CAILHAVA d'Estendoux, Jean François, 1739-1813.
—Le souper des petits-maîtres; conte composé de milie et un contes, par Cailhava de l'Estendoux. Illustrations de Pierre Gandon. Paris, Le Livre du Bibliophile [1956?] 2 p. l., 206 [2] p. col. front., 11 col. plates. 15 cm. (Le coffret du bibliophile illustré) I. Gandon, Pierre, illus. 2182948. Rare Books Dept.—B

CAILHAVA, Léon.
—De tristibvs Franciae libri qvatvor. 1840. *See* De tristibus Galliae. *PA 8485 D65 1840—L

Acquisition Costs

Table 10 details the full costs of this file to Minnesota. This file may be acquired for $80,000 purchase. There will be an alternate source of this file in the coming year as Blackwell-North America has acquired it and will be modifying its records for California. At this time Blackwell-North America has quoted prices of between $.10 - $.25 per record depending upon volume, method of selection, and the type of source record. For a probable 232,500 titles at $.25 the lowest acquisition cost total would be $58,125 versus the $80,000 purchase. Obviously this latter is a very unreliable figure. Therefore we believe use of this file for the 1975 period would have to be determined on the $80,000 direct purchase or $.106 per record figure. At a later period any new cost rate from Blackwell could be used in our cost calculations to replace these current figures as well as determine the extent of record improvements should the passage of time change the desirability of usage of foreign files.

Programming Costs

Because of the aforementioned complexities of this file it is assumed that major alterations via program would be needed, even to construct an L.C. Card No. index as well as specifically identify certain data fields distinctly from others, i.e. Place, publisher, and date would have to be scanned separately to identify them within the imprint tag.

One man year of programming is estimated to be required equalling $19,200. System utilization time would be approximately five times that for the NYPL file or $3,500. This would total $22,700 for programming/loading this file to usable state on our system.

In addition the printed catalogs would need to be used to obtain the California record identification numbers for some titles that result in incorrect index entries.

Temporary Storage Costs

This file is estimated to require approximately 525 million bytes of storage, i.e. 7 RJP04 86 megabyte disc drives and 7/8 ths of a controller. As this equipment could be used again for the last 30 months of its expected usable life for other purposes, one half of this cost is our true cost to store this data on the system for the project period exclusive of $.03 per record continued maintenance cost for the last 18 months of the period. This is one half of $186,375 or $93,188 to store this file for the project period.

Table 9. Costs of the University of California, Berkeyley, Five Year Union Catalog File for use on the University of Minnesota system for retrospective conversion.

Usable Records (All Non-MARC)	232,500	
Acquisitions costs		$ 80,000
Programming costs		$ 22,700
Temporary storage costs		$ 93,188
Total costs		$195,888

Cost per usable record	$.84
Costs exclusive of acquisitions costs	$115,888
Cost per usable record exclusive of acquisitions costs.	$.49

University of Chicago Library Catalog File

This file contains 250,000 records essentially covering the period 1968-to date. These records have been built from L.C. MARC II records, L.C. supplied copy and original University of Chicago cataloging during the period. The format of the record in the new system is MARC and uses MARC tags. The historical file uses differing tags than MARC but each data element can be defined in terms of a specific MARC II TAG and subfield. The University of Chicago is now completing work on converting this file to employ MARC II tag structure for their own system use.

The file was sampled for comparison through use of selected sample file listings which were sampled in turn to produce a random sample of the file.

Usable Records and Quality

This file produced a 38% total overlap with the Minnesota catalog. Of this 51% were MARC II derived. 43% are L.C. source cataloging copy and 6% were original copy. These percentages produce the following usable records:

Total 38%	95,000 records
51% MARC II	48,450 records
43% L.C. Source	40,850 records
6% Original	5,700 records

If again it is assumed that an existing MARC II file is on our system we would derive approximately 46,550 additional usable records from this file. About 49% of Minnesota's anticipated MARC II retrospective records could be obtained via this file. Assuming no presence of a MARC II file all 95,000 overlapping records would apply.

The quality of this file appears to be excellent, certainly on a par with NYPL file. As this investigator used a file dump and did not have any catalog card samples no sample data has been included here.

Acquisition Costs

The University of Chicago has not quoted the terms under which it would make its file available. Therefore, on a comparable basis to the NYPL file we could hypothesize a cost of $.106 per record or $26,500. Table 10. details the costs of this file to Minnesota.

Programming Costs

An estimate comparable to the costs for the NYPL file has been determined. The same 4 man months at $1,600 per month or $6,400 labor would apply. Due to the slightly larger file size we estimate $800 worth of system utilization

time would be required. The total would be $7,200.

Temporary Storage Costs

This file would require approximately 175 million bytes of disc storage, i.e. two 86 megabyte RJP04 units. This would bring the storage costs for this file to be roughly the same as for the NYPL file, $26,625 exclusive of the $.03 maintenance cost for the fiscal 18 months of the projected conversion period.

Table 10. Costs of the University of Chicago Library Catalog file for use on the University of Minnesota system for retrospective conversion.

Usable records total	95,000		
Usable Non-MARC II	46,550		
Usable MARC II	48,450		
Acquisitions costs		$ 26,500	
Programming costs		$ 7,200	
Temporary storage costs		$ 26,625	
Total costs		$ 60,325	
Costs per non-MARC usable record			$1.29
Costs per total usable records			$ .63
Costs exclusive of acqusitions costs			$33,825 total
Costs per non-MARC usable record exclusive of acquisition costs			$ .73
Costs per total usable records exclusive of acquisition costs			$ .36

Appendix 4. Supplementary information concerning in-house catalog record conversion as found in Section 6.

Possible conversion methods

There are many ways in which partial conversion of a library catalog has been accomplished. It may be that some service units or collections should be completely converted, i.e. such as an active course reserve collection. However, since 80% of the total book collection of the University is serviced via either the Wilson Library or subject branch libraries we should consider methods which identify the active portions of these collections for conversion.

Libraries have used circulation records to determine which titles to convert. They have also used publication date, language, service location, and subject. The most desirable way to determine which records to convert will be the way that achieves the greatest percentage of active titles at the least cost.

Selecting titles from circulation records

In this method for a large system it would be necessary to photocopy each

circulation card in the departmental file, alphabetize the cards, and search the union catalog for the main entry cards, perhaps photocopy them and refile them. After selection the record conversion process would proceed as described in Section 6. In a large catalog it would be difficult to remove drawers for extended periods to keyboard directly from the drawer itself if eight operators were doing the work. In this method as opposed to using a publication date cutoff we involve more steps and therefore incur a higher cost. Therefore, the question is really "Can a publication date cutoff achieve an equally high percentage of active titles as using the circulation records themselves?"

This investigator believes that the considerations discussed here show an excellent probability that use of publication date cutoffs can be valid for the University of Minnesota Libraries.

An examination of loans in various service units made in the Fall of 1974 revealed for example in the Wilson Library a range of 85,000 - 110,000 regular loans in the file at a given time. If such a file were taken at a point in time this number would result for conversion. Examination of other service units reveals a similar range of loans. With the known annual total circulation of the various units a ratio of maximal and minimal loans to total annual circulation was calculated. These values are 3.46 to 9.79. Dividing the total system charges of 1,250,104 by these values gives a maximum of 302,498 to 107,263 active titles. The minimum number is true if all repeat charges are for titles initially circulating at the beginning of the statistics period each year. Only if each file could be photocopied at its peak volume would the maximum number be identified. Thus the volume of active titles lies somewhere between these numbers.

The shelf list sample shows that a 1960 publication date cutoff would produce 377,500 titles. It appears that some number of titles approximate to the number circulating can be achieved through publication date cutoff. If this date is carefully chosen, the saving in labor to bypass the circulation record and move directly to the source record will be significant as the following costs have been determined:

A. Photocopy charge cards	
xerox labor	$2.70 per 200 titles
Xerox	$2.04 per 200 titles
Cutting and alphabetizing	$2.70 per 200 titles
B. Compare charge cards to card catalog and select main entry cards	
Labor	$4.05 per 200 titles
xerox labor	$2.70 per 200 titles
xerox	$2.04 per 200 titles
Total cost per 200	$16.23
Per title cost	.081
Cost for 300,000 titles	$24,300

Due to the partial nature of many of the shelf list cards in the University's file this alternative as a source document has been ruled out as even more costly than the circulation record method above.

The above cost of $24,300 would pay for the conversion of 13,500 titles for the system - not an insignificant amount.

Consequently the recommendation is that the catalog main entry card chosen by date of publication will achieve approximately the same number of titles for conversion and that circulating titles are primarily recent works in English and the more popular European languages, i.e. those that would occur if a wise publication date cutoff were used. For the system this date would appear to be from 1960 to 1965 with the earlier date perhaps better.

Using the Minicomputer System and CRT Terminals for Conversion

The physical conversion process of the titles once they have been selected covers:

Personnel training/documentation

Supervision, editorial, and error checking

Entry of data on the CRT terminals including
modification of an existing MARC record,
complete keyboard entry of a record
Terminal/communications hardware

Software costs

It is assumed that this process would occur using the PDP 11/40 on-line minicomputer equipped with sufficient terminals, file storage medium, MARC II data file, and data management system. The data management system as it has been developed would permit structuring the necessary indexes to retrieve existing records for modification, do any format conversions needed for entering into the permanent catalog file, and perform text modifications of the records, entered or identified in the MARC file. Moreover, the editing and proofreading function could be run in parallel through personnel using similar terminals calling forth the previously modified or originally entered record.

In this method there would be no intermediate document such as a coding form. Editors would scan the identified main entry cards prior to their handling by a terminal operator. During this scan, editors would label any ambiguous data fields or infrequently encountered structure, then terminal operators would input or search from those cards, entering or modifying records as required. Finally cards would be handed to other terminal operators who would call up these records and perform a visual check of the data and fields, correcting as required. This would end the conversion process, as no recataloging activity is included here. Any inconsistencies or redundancies present would be those in the existing manual catalog or from the MARC II or other pre-machine readable cataloging records. Cleaning of cataloging inconsistencies and redundancies are considered to be part of the normal usage of the on-line system files and not part of the conversion. Entry of copy information also has not been included here as this would have to be done from the shelf list and most likely incidental to an inventory of the shelves. This would be included as part of circulation system initiation costs rather than a basic bibliographic file conversion.

Personnel Training/Documentation

Three months will be needed to recruit and train the conversion staff (about 6 weeks actual training). Moreover, the Project Director will require three months of lead time to prepare training aids and documentation as well as accomplish loading of the MARC II file (if not already done) and any other file required. The following costs are associated with this area:

1. Project Director - 6 months at 1,600 month	$ 9,600
2. Documentation preparation	1,000
3. Personnel training expenses	
CRT operators - 16 at 1,170 each 6 weeks	18,720
Editors - 4 at 1,689	6,756
	$36,076

Per record cost of $.095

Supervision, Editorial, Error Checking

The costs of this over the 30 month project begin at the end of the six months of initial organization and training for the project. These are:

Project Director - 30 months at $1,600	$48,000
Editors - (4) x 30 months equals 120 months at $1,126 per month	135,120
CRT operators proofreading - (8) x 30 months equals 240 months at 780 per month	187,200
	$370,320

Per record cost of $.98

In this estimate it has been assumed that the production rate of the CRT operators proofreading will be the same as those inputting since a fatigue of looking at the screen will affect the rate of production. Probably in practice this portion will actually proceed slightly faster than has been estimated here, but from experience at this time there is no certainty of this.

Terminal Data Entry

In this phase the main entry cards checked by the editors will be entered into the system. The presence of the MARC file is assumed against which C. card number and author/title searches can be made. The operator would rch the specific title and if found modify the record inputting the University Minnesota classification number, the book's location, any locally supplied es or other data and then note our system I.D. number on the card and hand it to a proofreading CRT operator. This operator would proceed to call the record up for review. The terminal data entry costs are assessed on the basis of an average of 75 records per day input by each operator. It can be

assumed that approximately 100,000 of the titles would be in premachine readable form if the MARC file were present on the system--so this rate may be low as we established this rate through sending/keyboard of full bibliographic screens on the Bio-Medical Library PDP 11/40 with Super Bee SB-1 terminals. This is about a 35 wpm keying rate. To this must be added the cost of the terminals--in this case the purchase would equal the lease over 30 months so we have figured the full purchase price plus an estimated maintenance cost per terminal per year. Communications equipment is also included in the hardware figure. Therefore these costs are:

CRT input operators (8) x 30 months equals 240 months at 780 per month	$187,200
16 SuperBee SB-1 terminals at $3,300 each plus communications multiplexer, line adaptors, lines at $10,000	$ 62,800
Equipment maintenance over 30 months	$ 5,000
Total cost	$255,000

Per record cost of $.68

It has not been possible to determine exactly how much time is saved by modifying an existing machine readable record. However, from experience to date our assumption is that 30% faster input would result as the record must be searched, then every field scanned to determine the modification required, and finally the actual entry of the modifications. Therefore, if this is true the per record cost in the input stage could be lowered by only $.20 or to about $.48 per record for the operator portion of the cost. If 100,000 such records exist there is a potential offset of $20,000 on the conversion of a file on this basis.